

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: September 17, 2005, 08:52:14 ; Search time 178 Seconds
(without alignments)
1556.376 Million cell updates/sec

Title: US-10-615-659-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTAEGALRPPPGCKGS 541

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1753.5	61.4	781	2 Q8CON7	Q8CON7 mus musculus
2	1726.5	60.5	518	2 Q8C0V2	Q8C0V2 mus musculus
3	1052	36.9	704	2 Q8BV51	Q8BV51 mus musculus
4	1043.5	36.6	744	2 Q9H876	Q9H876 homo sapien
5	981.5	34.4	434	2 Q8NDN8	Q8NDN8 homo sapien
6	957.5	33.5	352	2 Q6AWA3	Q6AWA3 homo sapien
7	956.5	33.5	352	1 TTL3 HUMAN	Q9Y477 homo sapien
8	834	29.2	286	2 Q922T0	Q922T0 mus musculus
9	746.5	26.2	580	2 Q7Q156	Q7Q156 anopheles g
10	722.5	25.3	992	2 Q9VM91	Q9VM91 drosophila
11	536.5	18.8	719	2 Q9VM92	Q9VM92 drosophila
12	524.5	18.4	281	2 Q6ZU95	Q6ZU95 homo sapien
13	385	13.5	101	2 Q96G98	Q96G98 homo sapien
14	345.5	12.1	461	2 Q641W7	Q641W7 rattus norv
15	345	12.1	464	2 Q9D570	Q9D570 mus musculus
16	335	11.7	1075	2 Q7QT13	Q7QT13 giardia lam
17	331.5	11.6	794	2 Q6BFH6	Q6BFH6 paramecium
18	328	11.5	496	2 Q961I9	Q961I9 drosophila
19	328	11.5	496	2 Q9VX74	Q9VX74 drosophila
20	314	11.0	403	2 Q7Q2C8	Q7Q2C8 giardia lam
21	311	10.9	673	2 Q6ZV70	Q6ZV70 homo sapien
22	307	10.8	559	2 Q8CFV5	Q8CFV5 mus musculus
23	307	10.8	1333	2 Q8CHB8	Q8CHB8 mus musculus
24	303.5	10.6	1048	2 Q9PLV5	Q9PLV5 homo sapien
25	303.5	10.6	1226	2 Q9UPZ4	Q9UPZ4 homo sapien
26	303.5	10.6	1277	2 Q6EMB2	Q6EMB2 homo sapien
27	301	10.5	438	2 Q8C125	Q8C125 mus musculus
28	299.5	10.5	423	1 TTL1 HUMAN	Q95922 homo sapien
29	298.5	10.5	423	1 TTL1 MOUSE	Q91V51 mus musculus
30	297.5	10.4	827	2 Q9VKI9	Q9VKI9 drosophila
31	297.5	10.4	828	2 Q8IGW4	Q8IGW4 drosophila

RESULT 1

ID	Q8CON7	PRELIMINARY;	PRT;	781 AA.
AC	Q8CON7;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933401B17 product:hypothetical Tubulin-tyrosine ligase containing protein, full insert sequence.			
GN	Name=1700019P01Rik;			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;			
RX	Carninci P., Hayashizaki Y.;			
RT	"High-efficiency full-length cDNA cloning.";			
RL	Meth. Enzymol. 303:19-44(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;			
RA	RIKEN FANTOM Consortium;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";			
RL	Nature 420:563-573 (2002).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;			
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;			
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";			
RL	Genome Res. 10:1617-1630(2000).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;			
RA	Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto S., Ikegami T., Kashiwagi K.,			

32	297.5	10.4	1070	2	Q8IBB2	O8ipb2 drosophila
33	297.5	10.4	1295	2	Q6EEF3	Q6eeF3 cercopithec
34	292.5	10.2	440	2	Q7RKT2	Q7rkt2 plasmodium
35	291	10.2	541	2	Q8T417	Q8t417 drosophila
36	291	10.2	982	2	Q960F9	Q960f9 drosophila
37	291	10.2	989	2	Q9QV6	Q9qv6 drosophila
38	285.5	10.0	310	2	Q8I3W7	Q8i3w7 plasmodium
39	285	10.0	661	2	Q7QVA6	Q7qva6 giardia lam
40	282.5	9.9	487	2	Q9V297	Q9v297 drosophila
41	280.5	9.8	827	2	Q7PVY2	Q7pvY2 anopheles g
42	271.5	9.5	592	1	TTL2 HUMAN	Q9bw77 homo sapien
43	270	9.5	917	2	Q9VXK5	Q9vbx5 drosophila
44	269	9.4	1060	2	Q7QAD8	Q7qad8 anopheles g
45	265.5	9.3	436	2	Q7QZF0	Q7qzf0 giardia lam

ALIGNMENTS

```
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [16]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishii K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh K., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK030151; BAC26811.1; -
DR MGD; MGI:1922902; 1700019P01Rik.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR000276; GPCR Modופן.
DR InterPro; IPR004344; Tub_Tyr_lygase.
DR Pfam; PF03133; TTU; 1.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR_1; UNKNOWN_1.
DR Hypothetical protein; Ligase.
KW SEQUENCE 781 AA; 89397 MW; 9E98793C3351C3DE CRC64;
SQ
Query Match 61.4%; Score 1753.5; DB 2; Length 781;
Best Local Similarity 63.0%; Pred. No. 2.7e-112;
Matches 345; Conservative 69; Mismatches 113; Indels 21; Gaps 8;
QY 1 MASSILKVVVSHQS-CRSRSRSPRQREAEAGSSDLSSRQDAENABAKRLGLPGQLVDIA 59
DB 206 VAASILKVVVSHQNYC---SKVKGSKKEEAKNSDPSPKKDPENPLKPLSLSGQVVDTA 262
QY 60 CKVQCYALQGLEHEDIDTSADAVEDITEAEWEDLTQYYVSLVHGDAFINSRNYFSQCA 119
DB 263 CKVQCYALQGLEHEDIDVSEASTEALSEEWNDLTQYYVLLVHGNAISITDSKSYFAQCA 322
QY 120 LLNRITSVNPOTDIDGLRNIIKPAKSRGRDIVCMRVEETILEAADHPLSRDNKV 179
DB 323 LLSKISSVNPQTEIDGIRNIWIIPAKSRGRDIVCMRVENILSLVAADSQTTKDNKV 382
QY 180 VQKYIETPLLICDTPKIDRQWFLVTDNPLTITWFYKESYLRFSTQRFSLDKLDSAILHLCN 239
DB 383 VQKYIETPLMIYTKEDIRQWFLVTDNPLTITWFYKESYLRFSTQRFSLDKLDSAILHLCN 442
QY 240 NAVQYKNDVGRSPLLPANHMWSTTRFOEYIQRQGRGAVWGSVIYPSMKKIAHAMKVA 299
DB 443 NSIQRLKNDKERSPLLPCHNMTSTRFQEYLQKRGCGTWSIIYPSMKRAVTAAMRVA 502
QY 300 QDHVERKNSFELYGADFVLRGDFRPLWLEINSSPTMHPSPTPTAQLCAQVQEDTIKVAV 359
DB 503 QDHVEARKNSFELYGADFILGRDFKPLWLEINSSPTMHPSPTPTAQLCAQVQEDTIKVAV 562
QY 360 ----DRSCDIGNFELLWRQPVVEPPFPFSGSDLCVAGSVRRARRQVLPCVNLKASASLLD 415
DB 563 DRKLDNCIDIGNFELLWRQPAVELPFPNGSDLCVEGISVKKAKQMPPTASVGLSESLLD 622
QY 416 AQLKARGSPAMPDPAQGPSPALQDRDLGLKEKGLPLALLAPLRGAASGGGAQPTRYK 475
DB 623 APP-KVRSARALMETVIRPRTTRTQDWKREAEKVLSTTWSMPVMD-AEVRGRAKPIYAF 680
QY 476 AAGKVELPACPCHVDVSDQAPNTGVPAQKSWDNPQLNAHPLFVPL--RGLKTAEGALR 533
DB 681 EVNDYQ-----HVDNKGSKGSGTTRVQSSKV--PGVTLTSAQHAPALFAQTMRKPTQMTSS 731
```

```
QY 534 PPGGKGKS 541
DB 732 PPPTASGN 739
RESULT 2
QBCOV2
ID Q8COV2 PRELIMINARY; PRT; 518 AA.
AC Q8COV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4930524K07 product:hypothetical Tubulin-tyrosine ligase
DE containing protein, full insert sequence.
GN Name=1700019P01Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsuura H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Tanaka T., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
```


OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL613393; CAD38794.1;
 DR GO: GO:0004835; P: tubulin-tyrosine ligase activity; IEA.
 DR GO: GO:0006464; P: protein modification; IEA.
 DR InterPro: IPR004344; Tub_tyr_lygase.
 DR Pfam: PF01333; TTL; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 434 AA; 987956CA08651CAL CRC64;
 Query Match 34.4%; Score 981.5; DB 2; Length 434;
 Best Local Similarity 46.7%; Pred. No. 2.2e-59;
 Matches 194; Conservative 64; Mismatches 106; Indels 51; Gaps 5;
 QY 74 DIDSADAVEDL TEAEWEDLTQYYSLVHGDGAFISNSRNFYSCQALLNRITSVNPQTDI 133
 Db 2 DIDKDLAPLYLTPEGWSLFQRYQVVGEGALRHLDTVQRCEDILQQLQAVVPQIDM 61
 QY 134 DGLRNIIKPAKSRGRDVCMDRVEEILEAAADHPLSRDKNKVVQKVIETPLLICDT 193
 Db 62 EGRNIIWVPGAKSRGRGIMCHDHEMLKLVNGNPFVMDGKVVQKVIETPLIFGT 121
 QY 194 KPIRQWFLVTDNPLTIWFKSYRFRSTQSFSLDKLDSAIHLCNNAVQKLYKNDVGRS 253
 Db 122 KFDLRQWFLVTDNPLTIWFKSYRFRSTQSFSLDKLDSVHLNNSIQHLENSCHRH 181
 QY 254 PILPAHNMWTSRFOEVLQQRGAVGWSVIYSMKKAIHAHAKVAQDHVPEPRKNSFELY 313
 Db 182 PLLPPDNMSSQRFQALHOGMAGNAPNSTIIVPGMDAVIHALQTSQDTVQCRKASFELY 241
 QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAV-----DRSCDIGNFE 369
 Db 242 GADFVGEDFQPLWIEINASPSTAPSTAVTARLCAGVQADTLRVVIDRDLRNCDTGAFE 301
 QY 370 LLWRQPVVEPPFSGDLCVAGVSRRARRQVLPVNCNLKASALLDAQLPKRGPSAMPD 429
 Db 302 LIYKQPAVEPVQVGIKLLVEGFTIK-----PNAM-- 332
 QY 430 PAQGPSALQRLDGLKEEKGFLPALLAPLRGAESGGAQPTRTKAGKVELPA 484
 Db 333 -----CHRRMGVR-----PAVPLLTQRGSGEARHHPFSLHTKA-----QLPS 369
 RESULT 6
 Q6AWA3 PRELIMINARY; PRT; 352 AA.
 ID Q6AWA3
 AC Q6AWA3;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein DKFZp686D076.
 GN Name=DKFZp686D076;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervix;
 RA The German cDNA Consortium;
 RA Ansoorge W., Krieger S., Regiert T., Rittmuller C., Schwager B.,
 RA Mewes H.W., Weil B., Amid C., Oanger A., Fobo G., Han M., Wiemann S.;
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BX648175; CAH10554.1;
 DR GO: GO:0004835; F: tubulin-tyrosine ligase activity; IEA.
 DR GO: GO:0006464; P: protein modification; IEA.
 DR InterPro: IPR004344; Tub_tyr_lygase.
 DR Pfam: PF01333; TTL; 1.

KW Hypothetical protein.
 SQ SEQUENCE 352 AA; 40257 MW; 49FD8E9CAE1CB20D CRC64;
 Query Match 33.5%; Score 957.5; DB 2; Length 352;
 Best Local Similarity 54.2%; Pred. No. 7.4e-58;
 Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
 QY 74 DIDSADAVEDL TEAEWEDLTQYYSLVHGDGAFISNSRNFYSCQALLNRITSVNPQTDI 133
 Db 2 DIDKDLAPLYLTPEGWSLFQRYQVVGEGALRHLDTVQRCEDILQQLQAVVPQIDM 61
 QY 134 DGLRNIIKPAKSRGRDVCMDRVEEILEAAADHPLSRDKNKVVQKVIETPLLICDT 193
 Db 62 EGRNIIWVPGAKSRGRGIMCHDHEMLKLVNGNPFVMDGKVVQKVIETPLIFGT 121
 QY 194 KPIRQWFLVTDNPLTIWFKSYRFRSTQSFSLDKLDSAIHLCNNAVQKLYKNDVGRS 253
 Db 122 KFDLRQWFLVTDNPLTIWFKSYRFRSTQSFSLDKLDSVHLNNSIQHLENSCHRH 181
 QY 254 PILPAHNMWTSRFOEVLQQRGAVGWSVIYSMKKAIHAHAKVAQDHVPEPRKNSFELY 313
 Db 182 PLLPPDNMSSQRFQALHOGMAGNAPNSTIIVPGMDAVIHALQTSQDTVQCRKASFELY 241
 QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAV-----DRSCDIGNFE 369
 Db 242 GADFVGEDFQPLWIEINASPSTAPSTAVTARLCAGVQADTLRVVIDRDLRNCDTGAFE 301
 QY 370 LLWRQPVVEPPFSGDLCVAGVSRRARRQVLPVNCNLKASALLDAQLPKRGPSAMPD 429
 Db 302 LIYKQPAVEPVQVGIKLLVEGFTIK-----PNAM-- 332
 RESULT 7
 TTL3 HUMAN STANDARD; PRT; 352 AA.
 ID TTL3 HUMAN
 AC Q9Y4R7; Q9U199;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Tubulin tyrosine ligase-like protein 3 (HOTT1) (HQP0207).
 GN Name=TTL3;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 61-352 FROM N.A.
 RC TISSUE=Fetal liver;
 RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,
 RA Liu M., He F.;
 RT "Functional prediction of the coding sequences of 50 new genes deduced
 RT by analysis of cDNA clones from human fetal liver."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Contains 1 TTL domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AL096725; CAB46375.1;
 DR EMBL: AF078842; AAF23353.1;
 DR PIR: T12515; T12515.
 DR Genew: HGNC:24483; TTL3.
 DR InterPro: IPR004344; Tub_tyr_lygase.

```
DR Pfam; PF03133; TTL; 1.
KW Ligase.
FT DOMAIN 1 293
FT CONFLICT 67 67 I -> F (in Ref. 2).
SQ SEQUENCE 352 AA; 40356 MW; 49FD8E81187C720D CRC64;

Query Match 33.5%; Score 956.5; DB 1; Length 352;
Best Local Similarity 54.2%; Pred. No. 8.7e-58;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DDTSDADVEDLTEAWEDLTQOYYSVLVGDPAFISNRNYFSQCQALLNRITSVNPQTDI 133
DB 2 DDDKOLEALPLYLTPGMSFLQRYQVHGEAEHLRLDTQVQRCEDILQLOQAVVQIDM 61
QY 134 DGLRNIIWKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQVQKYEITPLLICDT 193
DB 62 EGRNIIWIKPAKSRGRGIMCMHLEMLKLVNGPVVWKGKVVQVQKYEITPLIFGT 121
QY 194 KFDIRQWFLVTDWNPITWIFYKESYLRFSTQFSLDKLDSAIHLCNNAVQKYLKNDVGRS 253
DB 122 KFDLRQWFLVTDWNPITWIFYKESYLRFSTQFSLDKLDSAIHLCNNVSHLNCNSIQKLENSCHRH 181
QY 254 PLLPAHNMWTSRFOYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 313
DB 182 PLLPPDNMWSRFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDVTQCRKASFELY 241
QY 314 GADFVLGRDFRPLWIEINSSPTMHPSTPTVTAOLCAQVQEDTIKVAV- ---DRSCDIGNFE 369
DB 242 GADFVEGEDEFQPLWIEINASPTWAPSTAVTARLCAGVQADTLRVVIDRDLRNCNCDTGAF 301
QY 370 LLMWQPWVEPPPPSGSDLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 8
Q922T0 PRELIMINARY; PRT; 266 AA.
AC Q922T0
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 4833441J24Rik protein.
GN Name=4833441J24Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Matra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006830; AAH06830.1; -.
DR MGD; MGI:2141418; 4833441J24Rik.
DR GO; GO:0004835; P:ubulin-tyrosine ligase activity; IEA.
DR GO; GO:0004644; P:protein modification; IEA.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 266 AA; 30506 MW; F8B8FB52FA8B8E98 CRC64;

Query Match 29.2%; Score 834; DB 2; Length 266;
Best Local Similarity 60.6%; Pred. No. 1.7e-49;
Matches 149; Conservative 44; Mismatches 49; Indels 4; Gaps 1;

QY 133 IDGLRNIIWKPAKSRGRDIVCMRVEIELEAAADHPLSRDNKVVQVQKYEITPLLICD 192
DB 1 MEGDQNIWIKVPGKSRGRGIMCMHLEMLKLVDCNPMKDGKVIQVQKYEITPLIFG 60
QY 193 TKFDIRQWFLVTDWNPITWIFYKESYLRFSTQFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
DB 61 TKFDIRQWFLVTDWNPITWIFYKESYLRFSTQFSLDKLDSAIHLCNNVSHLNCNSIQKLEASCHR 120
QY 253 SPLPAHNMWTSRFOYLRQGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
DB 121 HPMLPPDNMWSRFOAHLQEVDAKAWSSVIVPGMKAAVIHALQTSQDQNVQCRKASFEL 180
QY 313 YGADFVLGRDFRPLWIEINSSPTMHPSTPTVTAOLCAQVQEDTIKVAV- ---DRSCDIGNF 368
DB 181 YGADFVFGEDFQPLWIEINASPTWAPSTAVTARLCAGVQADTLRVVIDRDLRNCNCDTGAF 240
QY 369 ELLWRQ 374
DB 241 ELIYKQ 246

RESULT 9
Q7Q156 PRELIMINARY; PRT; 580 AA.
AC Q7Q156;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AGCP8623 (Fragment).
GN Name=agCG50567; ORFNames=ENSGG00000019848;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01006980; RAAL3905.1; -.
DR GO; GO:0004835; P:ubulin-tyrosine ligase activity; IEA.
DR GO; GO:0004644; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
FT NON_TER 1
FT NON_TER 580 580
SQ SEQUENCE 580 AA; 68072 MW; EE6C70E5FD9DB0C CRC64;

Query Match 26.2%; Score 746.5; DB 2; Length 580;
Best Local Similarity 39.8%; Pred. No. 5.2e-43;
Matches 159; Conservative 70; Mismatches 134; Indels 37; Gaps 9;

QY 5 ILKVVVSHQSCRSRSRSPRDQREAGSSDLSSRDQENAEAKRLPQLVVDIACKVCQ 64
DB 197 LLKWLIERHG-----TSGGGGL-----DAIADD--GNVPSTCITFALTRCK 237
```

```
QY 65 AYLGQLEHEDIDTSADAVEDLTAEWEDLTQOYSLVHGD---AFISNSR-----NYFS 115
Db 238 EYLDYCLHNDIDIEDT--KWDHWDVFTLHYLLTHEDNRIOQLKBEADAIEHYLA 295
QY 116 QCQALLNRITSVNPQTDIDGLRNIWIKPAKGRDVCMDRVEETLEAAADHPLSRD 175
Db 296 EAKSVLEQIKSHWPQYALDGYLNIWIKVPGNKGRCIGIHLNNIKOI--IAMVNPPIVSK 353
QY 176 NKWVQKIYITPLLICTKFDIRQWFLVTDWNPITWIFYKESYLRFSSTQFSLDKLDSAI 235
Db 354 TRYVIQKIERPLIHTKFDIRQWFMITSVQPLNIWIFYKESYLRFSSTQFSLDKLDSAI 413
QY 236 HLCNNVQKYLKNDVGSPLPAHNMWTSRFOYLQORQGVAGWSVVIYPMKKAIHA 295
Db 414 HLTHAIQKYYHNAV-RDEBLPHENMWDCHTFQAYLRFQIDKYEMWSERIYPGQKAIIGS 472
QY 296 MKVAQDHVEPRKNSFELYGADFVLGRDPRFWLWTEINSSPTMHPSTPVTALCAQVOEDTI 355
Db 473 LLACQDNMDRRPNTFFELYGADFMTEDFYFWLWTEINSSPDLASTSTVTLRCLQCQVEDTI 532
QY 356 KVAVDRSCD----IGNPELLWRQPVVEPPFPFSGDLCVAG 391
Db 533 RVVIDRRTDSNAPTSGFELYKQVIPKTPAYMGLNLQLRG 572
RESULT 10
Q9VM91 PRELIMINARY; PRT; 992 AA.
AC Q9VM91,
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE CG11323-PA.
OS ORFNames=CG11323;
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya S., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jallali M., Kalush K., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamo I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
```

```
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=22426065; PubMed=12537568;
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson C.R.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R.A., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
SEQUENCE FROM N.A.
RP MEDLINE=22426070; PubMed=12537573;
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
[4]
SEQUENCE FROM N.A.
RP MEDLINE=22426089; PubMed=12537572;
RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whittell E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
[5]
SEQUENCE FROM N.A.
RP FlyBase;
RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003614; AAF52432.1; -.
DR FlyBase; FBgn0031854; CG11323.
DR GO; GO:0004835; P:ubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_ty_lysase.
DR Pfam; PF03133; TTU; 1.
SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
Query Match 25.3%; Score 722.5; DB 2; Length 992;
Best Local Similarity 33.3%; Pred. No. 4.8e-41;
Matches 175; Conservative 86; Mismatches 198; Indels 67; Gaps 12;
QY 51 LFGQLVDIAKVCQAYLGQLEHEDIDTSADAVED---LTAEEWEDLTQOYSLVHGDFAFI 107
Db 326 IPVSADFAFKRLVEYIDSCQNDID----PEDPPKIWEHDWDAFLFQHQQLVNEGDRI 380
QY 108 SNS-----RNYFSQOALLNRITSVNPQTDIDGLRNIWIKPAKGRDVCMDRVEEI 162
Db 381 QHDGGQRLPEPMVKSLSLVDRMKVHPQYSLDGYQNMWIKVPAKGRGRIIIMDLKKI 440
QY 163 LELAAADHPLSRDKNKVVQKYYTETPLIICDTKFDIRQWFLVTDWNPITWIFYKESYLRFS 222
Db 441 --LGVNLSIAKSRIVVQKYYIERPLILFQTKFDIRQWFLVTDWNPITWIFYKESYLRFS 498
```



```
Db 353 MVQSTDLVLSNAKFKVSEKVMAYEARLAGLYDQIESLRDPYRWGDSRNLWILKPGYQSR 412
Qy 150 GRDIVCMRVEEIELEAAADHPLSRDNKVVVKIETPFLICDTKPIRQWFL--VTDWN 207
Db 413 GIGIVIRSSLDLILQWTSN---NQNKKYIVQKVIETPFLIYRTKPIRQVMLITITD-T 467
Qy 208 PLTIWFYKESYLFRESTQFSLDKLSDIAHLCCNNAVQKLVKNDVGRSPLLPAHNNMTSTRF 267
Db 468 KVSITWTRCYLRFSSQFTMDLRESIHLTNNSVQKRYKNTRDSRLPKNNNWSDQF 527
Qy 268 QEYLQROGRGAVGMSVITYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPL 327
Db 528 KNYLRINGAPDGSWSKTYNGFKQNLVAVWASLDETELLQNAFELYGDFMLDEHYNPIL 587
Qy 328 IEINSSPTMHPSTVTAQLCAQVEDIRKAVD-----RSCDIGNFELLWQPVVPPPPFS 383
Db 588 IEINSTPDLSPSTEITARICPMVLKDCIRVVVDLPKNPTAATGLFELAFE----- 637
Qy 384 GSDLCVAGSVRRARRQVLVPCNLKASALLDAQPLKARGPSA-----MPDPAQGPSPSA 438
Db 638 -----VNYISINK-----ADGPLELNGKQMTLFPNPMRNSPTRL 675
Qy 439 LQRLDLGLKEEGLPLALLAPLRGAESGAAPRTTKAAGKVELPACPCRHVDSPAQNTG 498
Db 676 LRKL-----NNVKTSTTKVE-----KVVERAPACNVK 703
Qy 499 VPVAQPAKS---WDPN 511
Db 704 NPTAKITKKVWYKN 718

RESULT: 12
Q6ZU95 PRELIMINARY; PRT; 281 AA.
AC Q6ZU95;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein FJ43887.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK125875; BAC6331.1; -.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 281 AA; 30962 MW; 3F4C4339709E92D0 CRC64;

Query Match 18.4%; Score 524.5; DB 2; Length 281;
Best Local Similarity 40.0%; Pred. No. 4e-28;
Matches 122; Conservative 34; Mismatches 76; Indels 73; Gaps 9;

Qy 233 SAHLCCNNAVQKLVKNDVGRSPLLPAHNNMTSTRFQYLQROGRGAVGMSVIVPSMKKAI 292
Db 17 SSVHLCCNNSIQKLENSCHRRHLLPPNNWSSQRFQHLQEMGAPNAWSTIIVPGMKDAV 76
Qy 293 AHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPMLEINSPTMHPSTVTAQLCAQVQE 352
Db 77 IHALQTSQDTVQCRKASFELYGADFVGFQFQWLEINASPTMAPSTAVTARLCAGVQA 136
Qy 353 DTTKAVV-----DRSCDIGNFELLWQPVVPPPPSGSDLCVAGSVRRARRQVLVPCNLK 408
```

```
Db 137 DTRVVIDRLDNCDTGAFELIYKQPAVEVPOVYVIRLLVEGFTIKK----- 184
Qy 409 ASASLLDAQPLKARGPSAMPDPAQ--PPSPAL-QRDLGLKEEKG--LPLALLAPLRGA 463
Db 185 -----PWAMCHRRMGVRPAVPLLTQRGSGEABVSGSLRKLPKVAQLR-RG 228
Qy 464 ESGGAQPTTKAAGKVELPAC-----PCRHVDSQAPNTGV 499
Db 229 TAGMQTQFVTISPA-STPRFSCLLPMYSYDTRARSSDDSTASWMLRPCR----- 276
Qy 500 PVAQP 504
Db 277 PQARP 281

RESULT 13
Q96GG8 PRELIMINARY; PRT; 101 AA.
AC Q96GG8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE TLL3 protein.
GN Name=TLL3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myer R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RL "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009479; AAH09479.1; -.
DR GO; GO:0004835; F.tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P.protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 101 AA; 12054 MW; 8EACD6C1131E0DF9 CRC64;

Query Match 13.5%; Score 385; DB 2; Length 101;
Best Local Similarity 68.0%; Pred. No. 4.3e-19;
Matches 68; Conservative 15; Mismatches 17; Indels 0; Gaps 0;

Qy 133 IGLRNIWIKPAKSRGRDIVCMRVEEIELEAAADHPLSRDNKVVVKIETPFLICD 192
Db 1 MEGDRNIWIKPAKSRGRGIMCMWHLKLVNGNVPVWKGKVVQKVIETPFLIFG 60
```


Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK015740; BAB29950.1; -;
DR MGD; MGI:1913589; 1700016P23Rik.
DR GO; GO:0004835; F:tubulin-cytosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_lyr_lygase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 464 AA; 53913 MW; 09FB48E438F433A4 CRC64;

Query Match 12.1%; Score 345; DB 2; Length 464;
Best Local Similarity 29.4%; Pred. No. 1.9e-15;
Matches 86; Conservative 52; Mismatches 95; Indels 60; Gaps 9;
Qy 140 WIIKPAKSRGRDIVCMRVEEILE-----LAAADHPLSRD 175
Db 146 WIMKPVARSQKGIPLRLRLKDIMWRKGTSGKKTGVTQPARANNPNSGSHDTRSSDD 205
Qy 176 NK-----WYVQKYIETPLICTFDIOWFLVTDWNPITWIFYKESYLRFSRFSL 228
Db 206 QKDDLPVENYVAQRYVENPYLIGRKFDLRVYVLVMSYIPLRWLYRDGPFARFSNTRFTL 265
Qy 229 DKLDS-AIHLNNAVOK-----YLNVDVGRSPLLPAHNMWTSTRFQYL-ORQGRGAVWG 281
Db 266 NSIDDHYVHLTNVAVQKTSPTYHLKGC-----KWWLQRFQYLAASKHGPKAV-- 313
Qy 282 SVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFLVGRDPRFWLIEINSSPTMHSTP 341
Db 314 ETLFSDMDNFIKLSQSVQKVIISDKHCFELYGYDILIDQLKPWLLVNASPSLTASSQ 373
Qy 342 VTAQLCAQVOEDTIKVA-----VDRSCDIGNFELLWRQPVVE----PPPFSG 384
Db 374 EDYELKTCLEDTLHVVDMEARLTGKEKRVGFDLMWMDGVPVSREDGSPDLSG 426

Search completed: September 17, 2005, 09:10:33
Job time : 181 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 08:51:23 ; Search time 172 Seconds
(without alignments)
1216.496 Million cell updates/sec

Title: US-10-615-659-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCRSRR.....LRGLKTAEGALRPPPGKGKS 541

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Genesecp16Dec04:*
- 2: Genesecp1980s:*
- 3: Genesecp1990s:*
- 4: Genesecp2000s:*
- 5: Genesecp2001s:*
- 6: Genesecp2002s:*
- 7: Genesecp2003as:*
- 8: Genesecp2003bs:*
- 9: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	541	8 ADJ93358	Adj93358 Human BGS
2	2843	99.6	541	8 ADJ93365	Adj93365 Human tub
3	1560	54.7	293	8 ADJ93366	Adj93366 Human BGS
4	1299	45.5	242	8 ADJ93360	Adj93360 Human BGS
5	1043.5	36.6	744	4 AAB94796	Aab94796 Human pro
6	1013	35.5	488	5 AAU74334	Aau74334 Human cyt
7	985.5	34.9	399	6 ABU11512	Abu11512 Human MDD
8	969.5	34.0	362	3 AAB58909	Aab58909 Breast an
9	956.5	33.5	352	3 AAB43005	Aab43005 Human ORF
10	956.5	33.5	352	4 AAM39450	Aam39450 Human pol
11	956.5	33.5	352	8 ADJ93457	Adj93457 Human HOT
12	956.5	33.5	352	8 ABM80420	Abm80420 Tumour-as
13	944.5	33.1	326	7 ADM05524	Adm05524 Human pro
14	865.5	30.3	292	8 ADJ93455	Adj93455 Human HOT
15	722.5	25.3	992	4 ABB65645	Abb65645 Drosophil
16	715	25.1	432	6 ABU92048	Abu92048 Human pro
17	536.5	18.8	719	4 ABB65541	Abb65541 Drosophil
18	524.5	18.4	281	8 ADQ66614	Adq66614 Novel hum
19	519	18.2	362	4 AAM41236	Aam41236 Human pol
20	375.5	13.2	566	4 ABG05971	Abg05971 Novel hum
21	336	11.8	160	8 ADH45424	Adh45424 Human mol
22	330	11.6	347	6 ABU00150	Abu00150 Human nov
23	328	11.5	496	4 ABB64074	Abb64074 Drosophil
24	311	10.9	673	8 ADQ66481	Adq66481 Novel hum
25	303.5	10.6	1281	4 AAM39105	Aam39105 Human pol

26	303.5	10.6	1304	4 AAM40891	Aam40891 Human pol
27	299.5	10.5	423	5 ABP43750	Abp43750 Ligase TT
28	299.5	10.5	423	8 ADJ93458	Adj93458 Human tub
29	299	10.5	393	7 ADC31411	Adc31411 Human nov
30	297.5	10.4	827	4 ABB60840	Abb60840 Drosophil
31	291	10.2	989	4 ABB60296	Abb60296 Drosophil
32	282.5	9.9	487	4 ABB59096	Abb59096 Drosophil
33	280.5	9.8	439	7 ADB65504	Adb65504 Human pro
34	278	9.7	321	7 ADC33066	Adc33066 Human nov
35	274	9.6	49	8 ADJ93373	Adj93373 Human BGS
36	271.5	9.5	519	4 AAM79305	Aam79305 Human pro
37	271.5	9.5	524	4 ABB22915	Abb22915 Protein #
38	271.5	9.5	592	7 ADB65217	Adb65217 Human pro
39	270	9.5	917	4 ABB64837	Abb64837 Drosophil
40	259	9.1	379	8 ADJ93456	Adj93456 Pig tubul
41	255.5	8.9	1138	8 ADJ71954	Adj71954 Human PMM
42	254.5	8.9	377	4 AAM79068	Aam79068 Human pro
43	254.5	8.9	377	5 ABB83472	Abb83472 Human cyt
44	254.5	8.9	377	7 ADE47756	Ade47756 Human NOV
45	254.5	8.9	377	8 ADJ79026	Adj79026 Human NOV

ALIGNMENTS

RESULT 1

ADJ93358
ID ADJ93358 standard; protein; 541 AA.

XX AC ADJ93358;

DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein sequence SeqID2.

XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-gen; gastrointestinal-gen;
KW osteoproteic; endocrine-gen; antiparkinsonian; antiarthritic; antiaesthetic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human.

XX OS Homo sapiens.

XX PN WO2004005487-A2.

XX PD 15-JAN-2004.

XX PF 09-JUL-2003; 2003WO-US021605.

XX PR 09-JUL-2002; 2002US-0394725P.

XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX PI Feder JN, Wu S, Nelson TC;

XX XX WPI; 2004-099381/10.

XX DR N-PSDB; ADJ93357.

XX PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX PT useful for preventing, treating or ameliorating a medical condition, e.g.
XX PT aberrant cellular proliferation, reproductive disorders or testicular
XX PT disorders.

PS Claim 5; SEQ ID NO 2; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary cancer,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC BGS-42 protein of the invention.

XX Sequence 541 AA;

Query Match 100.0%; Score 2854; DB 8; Length 541;
Best Local Similarity 100.0%; Pred. No. 3.6e-264;
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASSILKVVVSHOSCRSRKPRDREAGSGSDLSRQDAENAEAKRLGQLVDIAC 60
Db 1 MASSILKVVVSHOSCRSRKPRDREAGSGSDLSRQDAENAEAKRLGQLVDIAC 60
QY 61 KVCQAYLGQLEHEDIDTSADAVEDLTAEWEDLTQQYYSILVHGDAFISNRYFSQCQAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTAEWEDLTQQYYSILVHGDAFISNRYFSQCQAL 120
QY 121 LNRITSVNPQTDIDGLRNITWIKPAKSRGRDITVCMDRVVEEILEAAADHPLSRDNKVV 180
Db 121 LNRITSVNPQTDIDGLRNITWIKPAKSRGRDITVCMDRVVEEILEAAADHPLSRDNKVV 180
QY 181 OKYIETPLICDTPKFDIOWFLVTDWNPITWIFYKESYLRFSTQFSLDKLSAHLNWN 240
Db 181 OKYIETPLICDTPKFDIOWFLVTDWNPITWIFYKESYLRFSTQFSLDKLSAHLNWN 240
QY 241 AVQKYLKNDVGRSPLLPAAHNMWTSRFEYLQQRGAVMGSVIYPSMKKAIAHAMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLPAAHNMWTSRFEYLQQRGAVMGSVIYPSMKKAIAHAMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSTPVAQLCAQVQEDTIKAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHPSTPVAQLCAQVQEDTIKAVD 360
QY 361 RSCDIGNFELLWQPVVEPPPSGSDLCVAGYSVRARQVLPVCNLKASASILLDAQPLK 420
Db 361 RSCDIGNFELLWQPVVEPPPSGSDLCVAGYSVRARQVLPVCNLKASASILLDAQPLK 420
QY 421 ARGPSAMPDPAQPPSPALQDRLGLKEEGLPLALLAPLRGAESGGAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDRLGLKEEGLPLALLAPLRGAESGGAQPTRTKAAGKV 480
QY 481 ELPACPCRHVDSQAPNTGVVPAQPAKSWDPNQLNAHLEPVLRGLKTAGALRPPPGGKG 540
Db 481 ELPACPCRHVDSQAPNTGVVPAQPAKSWDPNQLNAHLEPVLRGLKTAGALRPPPGGKG 540
QY 541 S 541
Db 541 S 541

RESULT 2

ADJ93365

ID ADJ93365 standard; protein; 541 AA.

XX ADJ93365;

DT 06-MAY-2004 (first entry)

DE Human tubulin tyrosine ligase protein consensus sequence SeqID13.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human.

OS Homo sapiens.

OS Synthetic.

PN WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

DR N-PSDB; ADJ93364.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.

XX Example 4; SEQ ID NO 13; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS

PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.

XX Claim 8; SEQ ID NO 15921; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX Sequence 744 AA;

Query Match 36.6%; Score 1043.5; DB 4; Length 744;

Best Local Similarity 40.5%; Pred. No. 2.8e-90;

Matches 234; Conservative 85; Mismatches 192; Indels 67; Gaps 12;

QY 2 ASSTLKVVSHQSCRSRSPDRDREAGSSDLSSRQDAENAEAKLRLPGQLVDLACK 61
DB 151 ARNVKLIVKSE-----WKSPYIOAVEEASGDKPKKQKNPVL-----VSPFVDEALC 201
QY 62 VCQAYLQGLEHEDIDTSADAVEDLTEAWEDLTQYYLSLVHGDFAFISNRNYSQCOALL 121
DB 202 ACEEYLSNLAHMDIDKLEAPLYLTPGWSLFLQRYQYVQVHGAELRHLDTQVQRCEDIL 261
QY 122 NRITSVNPQTDIGLRNIWIIPKAASGRDVCMDRVEILLELAADHPLSRDNKKVVQ 181
DB 262 QQLQAVVPOIDMEGRNIWIIPKGAASGRGIMCMHLEMLKLVNGNPNVVMKDGKVVQ 321
QY 182 KYTETPLLICDTKFDIRQWFLVTDWNPPLTIWFKESYLRFSTQFSLDKLDSAIHLGNA 241
DB 322 KYTERPLLIIFGTGKPDLRQWFLVTDWNPPLTIWFKESYLRFSTQFSLDKLDSAIHLGNA 381
QY 242 VQKYLKNDVGRSPILLPAHNMWTSRFOYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQD 301
DB 382 IQHLENSCHRHPLPLPDNWSQQFQAHLEMGAPNAWSTIIVPGMKDAVIHALQTSQD 441
QY 302 HVEPRKNSFELYGADFVGRDFRPLWLEINSSPTMHPSTPTWALQCAQVEDTIKAV-- 359
DB 442 TVQCKKASFELYGADFVGFEDFQWLEINASPTMAPSTAVTARLCAVGQADTLRVVIDR 501
QY 360 --DRSCDIGNFELLWRQVPEPPFSGSDLCVAGSVRRARQVLPVCNLKASASLLDAQ 417
DB 502 MLDKNCDTGAFELIYKQPAVEVQYVGIIRLLVEGFTIK-----PMACHRRMGVRPAV 555
QY 418 PLKARGPS-----AMPDPAQPPSPALQDRLGL-----KEEGK 450
DB 556 PLLTQSGSGEARHHFSLHTKQ-LPSPHVLHOGQVLRQHSKLVGTKALSTTGKALRT 614
QY 451 LPLA-----LIAPLRGAESGAAPTAKA-----AGKVELPACPCRHVDQ--AP-NTG 498
DB 615 LPTAKVFISLPPNLDFFKVPASILKPKAPALLCLLRGPQLEVPVCCPLKSEQFLAPVGRS 674
QY 499 VPVAQPAKSWDPNQLNAPLEPVLRLGLKTAEGALRPPP 536

Db 675 RPKANSRPDCDKPRAEACPMKRL-----SPLKPLP 704

RESULT 6

AAU74334

ID AAU74334 standard; protein; 488 AA.

XX AAU74334;

XX 12-MAR-2002 (first entry)

XX Human cytoskeleton-associated protein (CYSKP) #5.

XX Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder;
XX cell proliferative disorder; inflammatory disorder; prion disease;
XX vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
XX neurological disorder; cell motility disorder; reproductive disorder;
XX spinal cord disease; central nervous system disorder; mental disorder;
XX gene therapy; cancer.

XX Homo sapiens.

XX WO2001185942-A2.

XX 15-NOV-2001.

XX 03-MAY-2001; 2001WO-US014355.

XX 05-MAY-2000; 2000US-0201960P.

XX 08-MAY-2000; 2000US-020729P.

XX 05-JUN-2000; 2000US-0209705P.

XX 07-JUN-2000; 2000US-0210149P.

XX 21-JUN-2000; 2000US-0213215P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;

XX Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;

XX Policky JL;

XX WPI; 2002-062248/08.

XX N-PSDB; AAS99894.

XX New cytoskeleton-associated proteins and polynucleotides, useful for

XX diagnosing, preventing and treating cell proliferative, autoimmune,

XX inflammatory, neurological, cell motility, reproductive and muscle

XX disorders.

XX Claim 1; Page 130-131; 194pp; English.

XX The invention relates to human cytoskeleton-associated polypeptides

XX (CYSKP) and their associated polynucleotide sequences. The sequences are

XX useful in the treatment of disorders associated with overexpression or

XX underexpression of CYSKP in a patient. The disorders include cell

XX proliferative disorders (such as cancer, actinic keratosis,

XX arteriosclerosis, cirrhosis, hepatitis and psoriasis),

XX autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,

XX osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus

XX and anaemia), vesicle trafficking disorders (such as

XX hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),

XX gastrointestinal disorders, prion diseases, neurological disorders (such

XX as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,

XX Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis

XX and other motor neuron disorders), cell motility disorders, reproductive

XX disorders (such as endometriosis and polycystic ovary syndrome), muscle

XX disorders (such as myocardiitis, migraine, hypertension, hypoglycaemia,

XX myocardial infarction, epilepsy and muscular dystrophy), spinal cord

XX diseases, central nervous system disorders (such as Down syndrome and

XX cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

XX Sequences AAU74330-AAU74363 represent human CYSKP of the invention

XX Sequence 488 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX		
DT	27-MAR-2001	(first entry)
XX		
DE	Breast and ovarian cancer associated antigen protein sequence SEQ ID 617.	
KW	Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;	
KW	nontuberc; neuroprotective; antiviral; antiallergic; hepatotropic;	
KW	antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;	
KW	antibacterial; antifungal; antiparasitic; cardiant; immune disorder;	
KW	Addison's disease; allergy; autoimmune haemolytic anaemia;	
KW	autoimmune thyroiditis; diabetes mellitus; Crohn's disease;	
KW	multiple sclerosis; rheumatoid arthritis; ulcerative colitis;	
KW	cardiovascular disorder; wound healing; neurological disease.	
OS	Homo sapiens.	
XX		
PX	WO200055173-AI.	
PD	21-SEP-2000.	
PF	08-MAR-2000; 2000WO-US005881.	
PR	12-MAR-1999; 99US-0124270P.	
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Ruben SM;	
PT	WPI; 2000-611515/58.	
DR	N-PSDB; AAF21812.	
XX		
PT	New human breast and ovarian cancer associated gene sequences and the	
PT	polypeptides encoded by these genes, useful in the prevention, treatment	
PT	and diagnosis of cancer, immune disorders, cardiovascular disorders and	
PT	neurological diseases.	
XX		
PS	Claim 11; Page 1056-1057; 1299pp; English.	
CC	Sequences AAF21614 - AAF22031 represent DNA sequences encoding human	
CC	proteins AAB58711 - AAB59128. The DNA and protein sequences are	
CC	associated with breast and ovarian cancer. Included in the invention are	
CC	sequences AAF2032 - AAF22040 and AAB59129 which are used in the	
CC	isolation and characterisation of the DNA and protein sequences of the	
CC	invention. The breast and ovarian cancer associated DNA, protein, agonist	
CC	or antagonist sequences exhibit cytotstatic; immunosuppressive; nootropic;	
CC	neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;	
CC	antiinflammatory; antitumor; vulnary; anticonvulsant; antibacterial;	
CC	antifungal; antiparasitic and cardiant activity. The polynucleotide and	
CC	protein sequences are used in the diagnosis of cancer, particularly	
CC	breast and ovarian cancer. The nucleic acid sequences, proteins, agonists	
CC	and agonists may also be used in the diagnosis, prevention and treatment	
CC	of immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;	
CC	cardiovascular disorders such as myocardial ischaemias; wound healing;	
CC	neurological diseases such as cerebral anoxia and epilepsy; and	
CC	infectious diseases	
XX		
SQ	Sequence 362 AA;	
	Query Match 34.0%; Score 969.5; DB 3; Length 362;	
	Best Local Similarity 52.0%; Pred. No. 1.le-83;	
	Matches 183; Conservative 58; Mismatches 92; Indels 19; Gaps 3;	
Qy	49 RGLPEQLVDIAKVCVAGLGHEDIDTSADAVDLTEAWEDLTQQYSLVHGDAFIS 108	
Dd	: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :	
Dd	9 RGVPYQ-----LGPHGRQ---GLEAPLYLTPEGWSLFQRYYQVHEGAEIR 53	
Qy	109 NSRNYFSOQALLNRITSVNPQTDDIGLRNLWTIKPKAKSRGRDTCMDRVEEILLEAAA 168	
Dd	::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :	
Dd	54 HLDTQVRCEDILOOLQAQVPOIDMEGRDNLIWKPGAKSRGRGIMCMDHLEEMLUKVNG 113	
Qy	169 DHPLSRDKNWVKRYETPLLCICTKFDIRFWFLTDWNPLTIWFYKESYLRFSTQSFSL 228	

Db	114	NPVYMGDKWVQYKIERPLLI	FTGKFDLRQWFLVTDNPLTVWFYRDSYIRFSTQPSL	173
Qy	229	DKLDSAIHLCHNNAVKYKLN	DVGRSPILLPAHNMVTSRFEYLRQGRGAVMGSVIYPSM	288
Db	174	KYLDNSVHLCHNNSIQK	HLNSCHRRPILLPPDNMSSQRFQAHLDWGA	PNASTIIVPGM 233
Qy	289	KKAIAHAMKVAQDHVEPR	KNSFELYGADFVLGRDPRPWLIEINSSPTVHPSTVTAQLCA	348
Db	234	KDAVIHALQTSQDTVQCR	KASFELYGADFVFGEDFQFWLIEINASTVAPSTAVTARLCA	293
Qy	349	QVQEDTTIKVAV----	DRSCDIGNFELLWRQPVREPPPPSGSDLCVAGSVRR	396
Db	294	GVQADTTLRWIDRXLD	RNCDTGAFLIYKQPAVEVPQVIGRLLEVGTIKK	345
RESULT 9				
AAB43005				
ID	AAB43005	standard; protein; 352 AA.		
XX	AC	AAB43005;		
XX	AC	AAB43005;		
DT	08-FEB-2001	(first entry)		
DE	XX	Human OREF ORF2769 polypeptide sequence SEQ ID NO:5538.		
XX	KW	Human; open reading frame; OREF; detection; cytostatic; hepatotropic;		
KW	KW	vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;		
KW	KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;		
KW	KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;		
KW	KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;		
KW	KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW	KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
KW	KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KW	KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;		
KW	KW	thrombosis; contraceptive.		
OS	XX	Homo sapiens.		
XX	XX	WO2000058473-A2.		
PN	XX	05-OCT-2000.		
PD	XX	31-MAR-2000; 2000WO-US008621.		
XX	XX	31-MAR-1999; 99US-0127607P.		
XX	XX	02-APR-1999; 99US-0127636P.		
PR	XX	05-APR-1999; 99US-0127728P.		
PR	XX	30-MAR-2000; 2000US-00540763.		
XX	XX	(CURA-) CURAGEN CORP.		
PA	XX	Shimkets RA, Leach M;		
PI	XX	WPI; 2000-602362/57.		
DR	XX	N-PSDB; AAC77214.		
DR	XX	Novel nucleic acids and peptides derived from open reading frame X,		
PT	XX	useful for treating e.g. cancers, proliferative disorders,		
PT	XX	neurodegenerative disorders and cardiovascular disease.		
XX	XX	Claim 11; Page 4720-4721; 5507pp; English.		
XX	XX	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,		
CC	XX	which represent the human OREF open reading frames 1 to 3161. The OREF		
CC	XX	sequences have activities such as: cytostatic; hepatotropic; vulnary;		
CC	XX	antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;		
CC	XX	anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;		
CC	XX	cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;		

CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 3; Length 352;
Best Local Similarity 54.2%; Pred. No. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNRNVFSQCALLNRITSVNPQTDI 133
DB 2 DIDKLEAPLYLTPEGWSLFLQRYQVQVHEGAEHLDTQVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKWWVQKYEIETPLLICDT 193
DB 62 EGDNRNIWIKPAKSRGRGIMCMCHLEEMKLNVGNPVVMKDGKWWVQKYEIETPLIFGT 121
QY 194 KFDIRQWFLVTDWNPITWIFKESYIRFSTQSFSLDKLSAHLCLNNAVQKYLKNDVGRS 253
DB 122 KFDLRQWFLVTDWNPITWIFRDSYIRFSTQSFSLKLNLSVHLNNSIQKHLNSCHRR 181
QY 254 PLLPAHNMWTSRFOEYLQRCRGAVGWSVYISMKKATAHAMKVAQDHVEPKNSFELY 313
DB 182 PLLPDDNMWSSQRFQAHLOEMGAPNAWSTIIVPGMKDAVIHALQTSQDITVQCKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAVV---DRSCDIGNFE 369
DB 242 GADFVFGEDFQFWLIEINASTPMASTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301
QY 370 LLWRQPVVEPPPSGSDLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 10
AAM39450
ID AAM39450 standard; protein; 352 AA.
XX
AC AAM39450;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2595.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58606.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2595; 10078pp; English.

PS The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification

XX Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. No. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNRNVFSQCALLNRITSVNPQTDI 133
DB 2 DIDKLEAPLYLTPEGWSLFLQRYQVQVHEGAEHLDTQVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIWIKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKWWVQKYEIETPLLICDT 193
DB 62 EGDNRNIWIKPAKSRGRGIMCMCHLEEMKLNVGNPVVMKDGKWWVQKYEIETPLIFGT 121
QY 194 KFDIRQWFLVTDWNPITWIFKESYIRFSTQSFSLDKLSAHLCLNNAVQKYLKNDVGRS 253
DB 122 KFDLRQWFLVTDWNPITWIFRDSYIRFSTQSFSLKLNLSVHLNNSIQKHLNSCHRR 181
QY 254 PLLPAHNMWTSRFOEYLQRCRGAVGWSVYISMKKATAHAMKVAQDHVEPKNSFELY 313
DB 182 PLLPDDNMWSSQRFQAHLOEMGAPNAWSTIIVPGMKDAVIHALQTSQDITVQCKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAVV---DRSCDIGNFE 369
DB 242 GADFVFGEDFQFWLIEINASTPMASTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301
QY 370 LLWRQPVVEPPPSGSDLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 11
ADJ93457
ID ADJ93457 standard; protein; 352 AA.
XX

AC XX
DT XX
XX XX
DE XX
XX XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory; Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiseborrheic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; HOTT13.
XX
OS Homo sapiens.
XX
XX WO2004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, e.g.
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Disclosure; SEQ ID NO 7; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiseborrheic or dermatological activity acting as
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC HOTT13 protein which is related to the invention. Note: The present
CC sequence does not appear in the specification but was obtained from
CC Genbank.
XX

SQ Sequence 352 AA;
Query Match 33.5%; Score 956.5; DB 8; Length 352;
Best Local Similarity 54.2%; Pred. No. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
Qy 74 DIDTSADAVDLTEAWEDLTQQYYSLVHGDAPFTSNRNYSFOCALLRITTSVNPQTDI 133
Db 2 DIDKDLAPLYLTPEGWSLFLQRYIQVHGEAUEHLDTVQRCEDILQQQAVVPQIDM 61
Qy 134 DGLRNIIKIPAAKSRGRDIVCMRVEEILELAADHPLSRDNKVVVQKYIETPLLICDT 193
Db 62 EGDRIWIVKPGAKSRGIGIMCDHLEMLKLVNGNPMVKDGKVVQKYIERPLIFGT 121
Qy 194 KFDIRQWFLVTDNPLTIWFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGRS 253
Db 122 KFDLRQWFLVTDNPLTIWFKESYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGRS 181
Qy 254 PLLPAHNMWTSRPOEYLOQGRGAVGVSIVYPSMKKAIAMKVAQDHVPRKNSFELY 313
Db 182 PLLPPDNWSSQRFQHLQEMGAPNAWSTIIVPMKDAVHALQTSQDTVQCRKASFELY 241
Qy 314 GADFLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
Db 242 GADFVGEDFQPLWLEINASPTWAPSTAVTARLCAGVQADTLRVVIDRMLDNCDDTGA 301
Qy 370 LLWRQPVVBPFPFSGSLCVAGV-SVRRAR 398
Db 302 LIYKQPVTTSPASTPRPSCLLPMYSDTRAR 331
RESULT 12
ABM80420
ID ABM80420 standard; protein; 352 AA.
XX
XX AC ABM80420;
XX
XX DT 18-NOV-2004 (first entry)
XX
XX DE Tumour-associated antigenic target (TAT) polypeptide PRO80956, SEQ.1058.
XX
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX
XX OS Homo sapiens.
XX
XX XX WO2004030615-A2.
XX
XX PD 15-APR-2004.
XX
XX PF 29-SEP-2003; 2003WO-US028547.
XX
XX PR 02-OCT-2002; 2002US-0414971P.
XX
XX PA (GSETH) GENENTECH INC.
XX
XX PI Wu TD, Zhang Z, Zhou Y;
XX
XX DR WPI; 2004-347921/32.
XX
XX DR N-PSDB; ACN37881.
XX
XX PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
XX PS Claim 12; SEQ ID NO 1058; 7273pp; English.
XX

CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX Sequence 352 AA;
SQ

Query Match 33.5%; Score 956.5; DB 8; Length 352;
Best Local Similarity 54.2%; Pred. No. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DITSDADVEDLTAEWEDLTQOYYSIVHGDAFISNRNYSFSCQALLNRITSVNPQTDI 133
DB 2 DIDKDLEAPLYLTPEGWSLFLQRIYQVHGEAELRLDTQVQCEDILQQLQAVVFOIDM 61
QY 134 DGLRNIIWIKPAKSRGRDIVCMRDVEEILELAADHPLSRDNKRWVVKYIETPLIICDT 193
DB 62 EGRDNIWIKPGAKSRGRGIMCNDHLEMLKLVNGPNVVKDGKVVVQKYEIRPLIIFGT 121
QY 194 KFDIROWFLVTDNPLTIWTFYKESYIRFSTQPSLKLDSAIHLCLNNAVKYIKNDVGRS 253
DB 122 KFDLQWFLVTDNPLTIWTFYRDSYIRFSTQPSLKLNSVHLCLNNSIQKHLNSCHRH 181
QY 254 PLLPAHNMWTSRFOELQQRGAVGWSVIVPSMKKATAHAMKVAODHVEPEKNSFELY 313
DB 182 PLLPPDNMSSQRFQHLQBMGAPNASTIIVPGMKDAVIHALQTSQDTVQCKKASFE 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
DB 242 GADFVFGEDFQPLWIEINASTPTWAPSTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301
QY 370 LLWRQPVPEPPPPSGSDLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCILLPMYSIDTRAR 331

RESULT 13

ADM05524

ID ADM05524 standard; protein; 326 AA.

XX

AC ADM05524;

XX

DT 20-MAY-2004 (first entry)

XX

DE Human protein of the invention SEQ ID NO:4209.

XX

KW human; gene therapy; diagnostic marker; pharmaceutical.

XX

OS Homo sapiens.

XX

PN EF1347046-A1.

XX

PD 24-SEP-2003.

XX

PF 12-APR-2002; 2002EP-00008400.

XX

PR 22-MAR-2002; 2002JP-00137785.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

DR WPI; 2003-723558/69.

DR N-PSDB; ADM03081.

XX New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.

XX Claim 1; SEQ ID NO 4209; 305pp; English.

XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC protein sequence of the invention.

XX Sequence 326 AA;

Query Match 33.1%; Score 944.5; DB 7; Length 326;
Best Local Similarity 54.3%; Pred. No. 2.4e-81;
Matches 176; Conservative 55; Mismatches 86; Indels 7; Gaps 2;

QY 74 DITSDADVEDLTAEWEDLTQOYYSIVHGDAFISNRNYSFSCQALLNRITSVNPQTDI 133
DB 2 DIDKDLEAPLYLTPEGWSLFLQRIYQVHGEAELRLDTQVQCEDILQQLQAVVFOIDM 61
QY 134 DGLRNIIWIKPAKSRGRDIVCMRDVEEILELAADHPLSRDNKRWVVKYIETPLIICDT 193
DB 62 EGRDNIWIKPGAKSRGRGIMCNDHLEMLKLVNGPNVVKDGKVVVQKYEIRPLIIFGT 121
QY 194 KFDIROWFLVTDNPLTIWTFYKESYIRFSTQPSLKLDSAIHLCLNNAVKYIKNDVGRS 253
DB 122 KFDLQWFLVTDNPLTIWTFYRDSYIRFSTQPSLKLNSVHLCLNNSIQKHLNSCHRH 181
QY 254 PLLPAHNMWTSRFOELQQRGAVGWSVIVPSMKKATAHAMKVAODHVEPEKNSFELY 313
DB 182 PLLPPDNMSSQRFQHLQBMGAPNASTIIVPGMKDAVIHALQTSQDTVQCKKASFE 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
DB 242 GADFVFGEDFQPLWIEINASTPTWAPSTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301
QY 370 LLWRQ---PVPEPPPPSGSDLCVA 390
DB 302 LIYKQGPAPNMQVSPERNAPLCPA 325

RESULT 14

ADJ93455

ID ADJ93455 standard; protein; 292 AA.

XX

AC ADJ93455;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human HOTT1 protein sequence SeqID2.

XX

KW testis-specific tubulin tyrosine-ligase-like polypeptide;

KW BGS-42 polypeptide; cytostatic; respiratory-gen; gastrointestinal-Gen;

KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;

KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antilasthmatic;

KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;


```

QY 108 SNS-----RNYFSQCOALLNRITSVNPQTDIDGLRNIWIIPAAKSRGRDIVCMRVEEI 162
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 381 QHGGQRLPEMVKSLVDKMKVHWPQYSLDGYQNMWIVKPANKCKRGIIILMDNLKKI 440
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 LELAAADHPLSRDNKVVQKIETPLLI CDTKEDIRQWFLVTDWNLTIWFYKESYLRF 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 --LGVVNLSTASKSRVYVQKVIERPLILFQTKEDIRQWFLIINTQPLVVMFYRESY_RFS 498
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 TORFSLDKLDSAIHLCNNAVQKYLKNDVGRSPLLPAHNMWTSTRFOEYLQROGRGAVWGS 282
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 499 SQEYSLSNHHESVHLTNVAILQKXYTNG-KDKKLPSENWDCYFQAYLRQIGKYNWMLE 557
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 VIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDPRFWLIEINSSPTMHPSTPV 342
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 558 RIFPGMRKAIVGCMLASQENMDRRPNTFELFGADFMCENFYFWLIEINSSPDLGATTSV 617
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 TAQLCAQVQEDTIKVAVDRSCD-----IGNFELLWRQPVVEPPFPFSGSDLCVAGSV- 394
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 TARMCPQCLEDEVVKVVIDRRTPKAEIKNFELAYRQVVPPTPAYMGLNLFVRGKQVLQKA 677
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 395 -----RRARROVLPCVNLKASASLIDAOPLKARGPSAMP----- 428
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 678 NHCGGHHYVYQOQRKERSLATSSVYRQRSAILHPATSI SRHRA MPTFNATEYMEKYNV 737
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 429 DPAQGPSPALQDLGLKEEKG LPLALLAPLRGAESGGAAPTRTYAAGKVELPACPCR 488
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 738 EPLSSRS-SLCSQLPKSPSAAPALTATP-----SGATSSYILKQAGRSITQLLSAT 789
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 489 H-----VDSQAPNTGVPAQPAKSWDPNQLNAHPLEPVLEGLK 526
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 790 HKRNTGGSLSGEGVQSTALP-PKRQSRGCGPRLSSTNPVESTKFKF 834
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: September 17, 2005, 09:07:30
Job time : 175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 23:48:14 ; Search time 698.935 Seconds
(without alignments)
8666.982 Million cell updates/sec

Title: US-10-615-659-1_COPY_369_1274
Perfect score: 906
Sequence: 1 gaggacatcgacgtcagc.....tcgagctctgtgaggcag 906

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq: *
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq: *
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq: *
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq: *
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq: *
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq: *
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq: *
22: /cgn2_6/ptodata/2/pubpna/US10J_PUBCOMB.seq: *
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq: *
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq: *
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	906	100.0	1838	19	US-10-615-659-1
2	906	100.0	1838	19	US-10-635-977-1
3	906	100.0	3554	19	US-10-615-659-12
4	906	100.0	3554	19	US-10-635-977-12
5	806.8	89.1	1859	19	US-10-615-659-10
6	806.8	89.1	1859	19	US-10-635-977-10
7	762	84.1	3465	19	US-10-615-659-11
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 12, Appl
					Sequence 12, Appl
					Sequence 10, Appl
					Sequence 10, Appl
					Sequence 11, Appl

8	762	84.1	3465	19	US-10-635-977-11	Sequence 11, Appli
9	760.4	83.9	1939	19	US-10-615-659-9	Sequence 9, Appli
10	760.4	83.9	1939	19	US-10-635-977-9	Sequence 9, Appli
11	726	80.1	726	19	US-10-615-659-3	Sequence 3, Appli
12	726	80.1	726	19	US-10-635-977-3	Sequence 3, Appli
c 13	427	47.1	490	16	US-10-029-386-24894	Sequence 24894, A
c 14	427	47.1	101270	20	US-10-723-860-631	Sequence 631, App
15	398.8	44.0	2380	18	US-10-275-595A-39	Sequence 39, Appl
16	398.8	44.0	2553	15	US-10-102-524-1765	Sequence 1765, Ap
17	398.8	44.0	2553	21	US-10-956-157-2016	Sequence 2016, Ap
18	398.8	44.0	2553	22	US-10-756-149-1971	Sequence 1971, Ap
19	398.8	44.0	2848	17	US-10-108-260A-1766	Sequence 1766, Ap
20	398.8	44.0	3001	15	US-10-037-270-494	Sequence 494, App
21	398.8	44.0	3001	17	US-10-117-722-494	Sequence 494, App
22	390.6	43.1	2611	9	US-09-925-298-199	Sequence 199, App
23	390.6	43.1	2611	14	US-10-102-806-199	Sequence 199, App
24	378.4	41.8	1400	21	US-10-956-157-9356	Sequence 9356, Ap
25	378.4	41.8	1673	21	US-10-956-157-4121	Sequence 4121, Ap
26	316.4	34.9	1400	21	US-10-956-157-7251	Sequence 7251, Ap
27	196	21.6	755	13	US-10-027-632-135929	Sequence 135929,
28	196	21.6	755	13	US-10-027-632-135930	Sequence 135930,
29	196	21.6	755	17	US-10-027-632-135929	Sequence 135929,
30	196	21.6	755	17	US-10-027-632-135930	Sequence 135930,
31	190.6	21.0	1958	21	US-10-956-157-4903	Sequence 4903, Ap
32	166	18.3	1728	22	US-10-450-763-5962	Sequence 5962, Ap
33	164.4	18.1	2250	22	US-10-450-763-5187	Sequence 5187, Ap
34	163.4	18.0	418	9	US-09-983-965-2032	Sequence 2032, Ap
c 35	130	14.3	525	16	US-10-029-386-11174	Sequence 11174, A
36	122.6	13.5	436	17	US-10-242-535A-7660	Sequence 7660, Ap
37	122.6	13.5	436	18	US-10-085-783A-7660	Sequence 7660, Ap
38	117	12.5	492	10	US-09-918-995-35253	Sequence 35253, A
39	113.4	12.5	279	18	US-10-424-599-119452	Sequence 119452,
40	92.2	10.2	207	9	US-09-728-445-450	Sequence 450, App
41	92.2	10.2	207	22	US-10-964-549-450	Sequence 450, App
c 42	80	8.8	80	19	US-10-615-659-52	Sequence 52, Appl
c 43	80	8.8	80	19	US-10-635-977-52	Sequence 52, Appl
c 44	58.2	6.4	1648	21	US-10-887-553A-838	Sequence 838, App
45	56.8	6.3	1080	15	US-10-156-761-2625	Sequence 2625, Ap

ALIGNMENTS

RESULT 1
US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: P0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Query Match 100.0%; Score 906; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 1.8e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGACATCGACACGTGACAGATGCGTGGAGACCTACTGAGCGCGAGTGGAGGAC 60
|||||

```
Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCAGTGGGAGGAC 428
QY 61 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCCAATTCAAGAAAT 120
Db 429 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCCAATTCAAGAAAT 488
QY 121 TACTTTTGGCAGTGCAGGCTCTGCTGTAATAGAAATCACGTCGTGTGAAACCCCTCAGACGGAC 180
Db 489 TACTTTTGGCAGTGCAGGCTCTGCTGTAATAGAAATCACGTCGTGTGAAACCCCTCAGACGGAC 548
QY 181 ATTGAGCGGCTCCGGACATCTGGATATATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
Db 549 ATTGAGCGGCTCCGGACATCTGGATATATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 608
QY 241 ATAGTGTGATGACACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACCAACCCCTCTT 300
Db 609 ATAGTGTGATGACACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACCAACCCCTCTT 668
QY 301 TCCAGGGACAACAAAGTGGGTGCTCCAGAAATGATCATCGAGACGCCGCTGCTCATCTGTGAC 360
Db 669 TCCAGGGACAACAAAGTGGGTGCTCCAGAAATGATCATCGAGACGCCGCTGCTCATCTGTGAC 728
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGAGAACCCCTGACCATCTGG 420
Db 729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGAGAACCCCTGACCATCTGG 788
QY 421 TTCTACAAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGTTCCTCCCTGGACAAGCTGGAC 480
Db 789 TTCTACAAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGTTCCTCCCTGGACAAGCTGGAC 848
QY 481 AGCGCCATCCACTGCTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCGC 540
Db 849 AGCGCCATCCACTGCTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCGC 908
QY 541 AGCCCCCTGCTGCCCGCACACAACATGTGGACACGACCAAGTTCCAGAGTACCTTGCAG 600
Db 909 AGCCCCCTGCTGCCCGCACACAACATGTGGACACGACCAAGTTCCAGAGTACCTTGCAG 968
QY 601 CGCCAGCGCATGAGTGGCCCGCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 720
Db 1029 CGCCAGCGCATGAGTGGCCCGCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 1088
QY 721 TACGGGGCTGACTTTCGTCCTTTGGGAGGAGCTTTCAGGCCCTCGGCTGATCGAGATCAATTCC 780
Db 1089 TACGGGGCTGACTTTCGTCCTTTGGGAGGAGCTTTCAGGCCCTCGGCTGATCGAGATCAATTCC 1148
QY 781 AGCCCCCATGACATGACCCCGTCCAGCGGTCACGCGGCTGATCGAGATCAATTCC 840
```

RESULT 2

```
US-10-635-977-1
; Sequence 1, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635, 977
```

```
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-635-977-1

Query Match 100.0%; Score 906; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 1.8e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCAGTGGGAGGAC 60
Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCAGTGGGAGGAC 428
QY 61 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCCAATTCAAGAAAT 120
Db 429 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCCAATTCAAGAAAT 488
QY 121 TACTTTTGGCAGTGCAGGCTCTGCTGTAATAGAAATCACGTCGTGTGAAACCCCTCAGACGGAC 180
Db 489 TACTTTTGGCAGTGCAGGCTCTGCTGTAATAGAAATCACGTCGTGTGAAACCCCTCAGACGGAC 548
QY 181 ATTGAGCGGCTCCGGACATCTGGATATATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
Db 549 ATTGAGCGGCTCCGGACATCTGGATATATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 608
QY 241 ATAGTGTGATGACACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACCAACCCCTCTT 300
Db 609 ATAGTGTGATGACACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACCAACCCCTCTT 668
QY 301 TCCAGGGACAACAAAGTGGGTGCTCCAGAAATGATCATCGAGACGCCGCTGCTCATCTGTGAC 360
Db 669 TCCAGGGACAACAAAGTGGGTGCTCCAGAAATGATCATCGAGACGCCGCTGCTCATCTGTGAC 728
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGAGAACCCCTGACCATCTGG 420
Db 729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGAGAACCCCTGACCATCTGG 788
QY 421 TTCTACAAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGTTCCTCCCTGGACAAGCTGGAC 480
Db 789 TTCTACAAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGTTCCTCCCTGGACAAGCTGGAC 848
QY 481 AGCGCCATCCACTGCTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCGC 540
Db 849 AGCGCCATCCACTGCTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCGC 908
QY 541 AGCCCCCTGCTGCCCGCACACAACATGTGGACACGACCAAGTTCCAGAGTACCTTGCAG 600
Db 909 AGCCCCCTGCTGCCCGCACACAACATGTGGACACGACCAAGTTCCAGAGTACCTTGCAG 968
QY 601 CGCCAGCGCATGAGTGGCCCGCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 720
Db 1029 CGCCAGCGCATGAGTGGCCCGCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 1088
QY 721 TACGGGGCTGACTTTCGTCCTTTGGGAGGAGCTTTCAGGCCCTCGGCTGATCGAGATCAATTCC 780
Db 1089 TACGGGGCTGACTTTCGTCCTTTGGGAGGAGCTTTCAGGCCCTCGGCTGATCGAGATCAATTCC 1148
QY 781 AGCCCCCATGACATGACCCCGTCCAGCGGTCACGCGGCTGATCGAGATCAATTCC 840
```

```
Db 1149 AGCCCCACCATGACCCGTCACGCGGTCACGCCCCAGCTGTGTGACAGGTGCAGGAG 1208
Qy 841 GACACCATCAAGTGGCGGTGGACCCGACGCTGTGACATCGGCACACTTCGAGCTCCTGTGG 900
Db 1209 GACACCATCAAGTGGCGGTGGACCCGACGCTGTGACATCGGCACACTTCGAGCTCCTGTGG 1268
Qy 901 AGGCAG 906
Db 1269 AGGCAG 1274

RESULT 3
US-10-615-659-12
; Sequence 12, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-12

Query Match 100.0%; Score 906; DB 19; Length 3554;
Best Local Similarity 100.0%; Pred. No. 2.1e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGCGGAGTGGAGGAC 60
Db 1911 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGCGGAGTGGAGGAC 1970
Qy 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAAT 120
Db 1971 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAAT 2030
Qy 121 TACTTTTCGAGTGCCAGGCTCTGCTGAATAGATACAGTCTGTGAACCCCTCAGACGGAC 180
Db 2031 TACTTTTCGAGTGCCAGGCTCTGCTGAATAGATACAGTCTGTGAACCCCTCAGACGGAC 2090
Qy 181 ATTGACGGCTCCGGAACATCTGGATTATAAAGCCCGGCGCAAGTCCCGGGGCGGAGAC 240
Db 2091 ATTGACGGCTCCGGAACATCTGGATTATAAAGCCCGGCGCAAGTCCCGGGGCGGAGAC 2150
Qy 241 ATAGTGTGCATGACCGCTGTGGAGGATCCTGGAGCTGCGAGCTGCGAGCTGCAGACCCCTCTT 300
Db 2151 ATAGTGTGCATGACCGCTGTGGAGGATCCTGGAGCTGCGAGCTGCGAGCTGCAGACCCCTCTT 2210
Qy 301 TCAGGGACAAACAGTGGGTGGTCCAGAAGTACATCGAGACGCGCTGCTCATCTGTGAC 360
Db 2211 TCAGGGACAAACAGTGGGTGGTCCAGAAGTACATCGAGACGCGCTGCTCATCTGTGAC 2270
Qy 361 ACCAAGTTCGACATCAGACAGTGTCTCGTCAAGGACTGGAACCCCTTGACCATCTGG 420
Db 2271 ACCAAGTTCGACATCAGACAGTGTCTCGTCAAGGACTGGAACCCCTTGACCATCTGG 2330
Qy 421 TTCTTACAGGAGGTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
Db 2331 TTCTTACAGGAGGTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 2390
Qy 481 AGGCCATCCACCTGTGCAACAGCCGCTCAGAAATGACATGATGTGGGCGCG 540
Db 2391 AGGCCATCCACCTGTGCAACAGCCGCTCAGAAATGACATGATGTGGGCGCG 2450
```

```
Qy 541 AGCCCCCTGCTGCCGACACATGTGGACCAAGTTCACAGGAGTACCTGCAG 600
Db 2451 AGCCCCCTGCTGCCGACACATGTGGACCAAGTTCACAGGAGTACCTGCAG 2510
Qy 601 CGCCAGGGCGGTGGCGGTGGTGGGCGAGCGTCACTTACCCGTCATGAAGAAGGCCATC 660
Db 2511 CGCCAGGGCGGTGGCGGTGGTGGGCGAGCGTCACTTACCCGTCATGAAGAAGGCCATC 2570
Qy 661 GCCCAGCCCATGAAGTGGCCAGGACCAAGTGGAGGCTCGCAAGACAGCTTTGAGCTC 720
Db 2571 GCCCAGCCCATGAAGTGGCCAGGACCAAGTGGAGGCTCGCAAGACAGCTTTGAGCTC 2630
Qy 721 TAGCGGCTGACTTCTGCTTGGGAGGACTTCAGGGCCCTGGCTGTATCGAGATCAATTC 780
Db 2631 TAGCGGCTGACTTCTGCTTGGGAGGACTTCAGGGCCCTGGCTGTATCGAGATCAATTC 2690
Qy 781 AGCCCCACCATGACCCGTCACGCCGTCACGCCGACAGTGTGTGCACAGGTGCAGGAG 840
Db 2691 AGCCCCACCATGACCCGTCACGCCGTCACGCCGACAGTGTGTGCACAGGTGCAGGAG 2750
Qy 841 GACACCATCAAGTGGCGGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
Db 2751 GACACCATCAAGTGGCGGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 2810
Qy 901 AGGCAG 906
Db 2811 AGGCAG 2816

RESULT 4
US-10-635-977-12
; Sequence 12, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-12

Query Match 100.0%; Score 906; DB 19; Length 3554;
Best Local Similarity 100.0%; Pred. No. 2.1e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGCGGAGTGGAGGAC 60
Db 1911 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGCGGAGTGGAGGAC 1970
Qy 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAAT 120
Db 1971 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAAT 2030
Qy 121 TACTTTTCGAGTGCCAGGCTCTGCTGAATAGATACAGTCTGTGAACCCCTCAGACGGAC 180
Db 2031 TACTTTTCGAGTGCCAGGCTCTGCTGAATAGATACAGTCTGTGAACCCCTCAGACGGAC 2090
Qy 181 ATTGACGGCTCCGGAACATCTGGATTATAAAGCCCGGCGCAAGTCCCGGGGCGGAGAC 240
Db 2091 ATTGACGGCTCCGGAACATCTGGATTATAAAGCCCGGCGCAAGTCCCGGGGCGGAGAC 2150
Qy 241 ATAGTGTGCATGACCGCTGTGGAGGATCCTGGAGCTGCGAGCTGCGAGCTGCAGACCCCTCTT 300
```

Db	2151	ATAGTGTGCATGGACCGGTGTGGAGGAGATCCTGGAGTGGCAGCTGCAGACCAACCCCTCTT	2210
Qy	301	TCCAGGACAACAAGTGGGTGCAGAAGTACATCGAGCGCGCTGCTCATCTGTGAC	360
Db	2211	TCCAGGACAACAAGTGGGTGCAGAAGTACATCGAGCGCGCTGCTCATCTGTGAC	2270
Qy	361	ACCAAGTTCGACATCAGACAGTGGTTCTCGTCA CGGACTGGAAACCCCTGACCATCTGG	420
Db	2271	ACCAAGTTCGACATCAGACAGTGGTTCTCGTCA CGGACTTGGAAACCCCTGACCATCTGG	2330
Qy	421	TTCTTCAAGAGAGTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGGCAAGCTGGAC	480
Db	2331	TTCTTCAAGAGAGTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGGCAAGCTGGAC	2390
Qy	481	AGGCCATCCACCTGTGCAACAACGCCGTCCAGAAGTACCTGAAAGATGATGTGGGCCGC	540
Db	2391	AGGCCATCCACCTGTGCAACAACGCCGTCCAGAAGTACCTGAAAGATGATGTGGGCCGC	2450
Qy	541	AGCCCCGTGTGCCCGCACACAATGTGSCACGACACCGAGTTTCCAGGAGTACCTGCGAC	600
Db	2451	AGCCCCGTGTGCCCGCACACAATGTGSCACGACACCGAGTTTCCAGGAGTACCTGCGAC	2510
Qy	601	CGCCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCACTTACCCTCCATGAAAGAGGCCATC	660
Db	2511	CGCCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCACTTACCCTCCATGAAAGAGGCCATC	2570
Qy	661	GCCCA GCCCATGAAGTGGCGCCAGGACACAGTGGAGCCTCGCAAGAACAGCTTTGAGCTC	720
Db	2571	GCCCA GCCCATGAAGTGGCGCCAGGACACAGTGGAGCCTCGCAAGAACAGCTTTGAGCTC	2630
Qy	721	TACGGGCTGACTTCGTCTTGGAGGGACTTACGGCCCTGGCTGTATCGAGATCAATTC	780
Db	2631	TACGGGCTGACTTCGTCTTGGAGGGACTTACGGCCCTGGCTGTATCGAGATCAATTC	2690
Qy	781	AGCCCCACCATGCNCCCGTCCAGCCCGGTCA CGGCCAGCTGTGTGCA CAGGTGCAGGAG	840
Db	2691	AGCCCCACCATGCNCCCGTCCAGCCCGGTCA CGGCCAGCTGTGTGCA CAGGTGCAGGAG	2750
Qy	841	GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	900
Db	2751	GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	2810
Qy	901	AGGCAG 906	
Db	2811	AGGCAG 2816	

```

RESULT 5
US-10-615-659-10
; Sequence 10, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG542
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-10

```

Qy	1	GAGGACATCGACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGGCCGAGTGGGAGGAC	60
Db	263	GAGGACATCGACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGGCCGAGTGGGAGGAC	322
Qy	61	CTGACCCAGCAGTACTACCTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	120
Db	323	CTGACCCAGCAGTACTACTCCCTCGTTTCATCTCCAATTCAAGAAAT	351
Qy	121	TACTTTTTCGAGTCGCCAGGCTCTGCTGAAATAGAAATCACGTCCTGTGAAACCCCTCAGACGGAC	180
Db	352	-----TCCAGGCTCTGCTGAAATAGAAATCACGTCCTGTGAAACCCCTCAGACGGAC	399
Qy	181	ATTGACGGGCTCCGGAAACATCTGGAATTATAAGCCCGCGGCCAAGTCCCGGGCGCGAGAC	240
Db	400	ATTGACGGGCTCCGGAAACATCTGGAATTATAAGCCCGCGGCCAAGTCCCGGGCGCGAGAC	459
Qy	241	ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT	300
Db	460	ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT	519
Qy	301	TCCAGGACCAACAGTGGGTGCACGAAAGTATACGAGACGCGCTGCTCATCTGTGAC	360
Db	520	TCCAGGACCAACAGTGGGTGCACGAAAGTATACGAGACGCGCTGCTCATCTGTGAC	579
Qy	361	ACCAAGTTCGACATCGACAGTGGTTCCTCGTCA CGGACTGGAACCCCTGACCACTCG	420
Db	580	ACCAAGTTCGACATCGACAGTGGTTCCTCGTCA CGGACTGGAACCCCTGACCACTCG	639
Qy	421	TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	480
Db	640	TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	699
Qy	481	AGGCCATCCACTGTGTCAACAAACGCGCTCCAGAAAGTACCTGAAGAATGATGTGGGCGCG	540
Db	700	AGGCCATCCACTGTGTCAACAAACGCGCTCCAGAAAGTACCTGAAGAATGATGTGGGCGCG	759
Qy	541	AGCCCCCTGTGCCCGCACACATGTGGAACAGACACAGGTTCCAGGAGTACTTCGCAG	600
Db	760	AGCCCCCTGTGCCCGCACACATGTGGAACAGACACAGGTTCCAGGAGTACTTCGCAG	819
Qy	601	CGCAGGGCGGTGGCGCGTGTGGGCGAGCTCATCTACCCGTCATGAAGAAGGCCATC	660
Db	820	CGCAGGGCGGTGGCGCGTGTGGGCGAGCTCATCTACCCGTCATGAAGAAGGCCATC	879
Qy	661	GCCACGCCCATGAAGTGGCCACGAGCACCGTGGAGCCTCGCAAGAACAGCTTTGAGCTC	720
Db	880	GCCACGCCCATGAAGTGGCCACGAGCACCGTGGAGCCTCGCAAGAACAGCTTTGAGCTC	939
Qy	721	TAGGGGCTGACTTCGTCTTGGGAGGGAATTAGGGCCCTGGCTGATCGAGATCAATTCC	780
Db	940	TAGGGGCTGACTTCGTCTTGGGAGGGAATTAGGGCCCTGGCTGATCGAGATCAATTCC	999
Qy	781	AGCCCCACCATGACCCGTCACCGCGTCCAGGCCCGACGCTGTGTGCACAGGTGCGAGGAG	840
Db	1000	AGCCCCACCATGACCCGTCACCGCGTCCAGGCCCGACGCTGTGTGCACAGGTGCGAGGAG	1059
Qy	841	GACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCCAATCTCGAGCTCTGTGG	900
Db	1060	GACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCCAATCTCGAGCTCTGTGG	1119
Qy	901	AGGCAG 906	
Db	1120	AGCGGG 1125	

RESULT 6
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG542

```
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10

Query Match      89.1%; Score 806.8; DB 19; Length 1859;
Best Local Similarity 95.0%; Pred. No. 8.7e-231;
Matches 861; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 1 GAGGACATCGACAGCTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGAC 60
DB 263 GAGGACATCGACAGCTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGAC 322
QY 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGCGGATGCTTTTCATCTCCAATTCAGAGAAAT 120
DB 323 CTGACCCAGCAGTACTACTCCCTCGTTTCATCTCCAATTCAGAGAAAT 351
QY 121 TACTTTTCGAGTGCAGAGCTCTGCTGAATAGAAATCACTGCTGTGAACCCCTCAGACGGAC 180
DB 352 -----TCCAGAGGCTCTGCTGAATAGAAATCACTGCTGTGAACCCCTCAGACGGAC 399
QY 181 ATTGAGCGGCTCCGGACATCTGGATTATAAAGCCGCGCGCAAGTCCCGGGGCCGAGAC 240
DB 400 ATTGAGCGGCTCCGGACATCTGGATTATAAAGCCGCGCGCAAGTCCCGGGGCCGAGAC 459
QY 241 ATAGTGTGCAATGGAACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT 300
DB 460 ATAGTGTGCAATGGAACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT 519
QY 301 TCCAGGAGCAACAAGTGGGTGTCGAGAGTAATCGAGAGCGCGTGTCTCATCTGTGAC 360
DB 520 TCCAGGAGCAACAAGTGGGTGTCGAGAGTAATCGAGAGCGCGTGTCTCATCTGTGAC 579
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGAGTACATCGAGAGCGCGTGTCTCATCTGTG 420
DB 580 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGAGTACATCGAGAGCGCGTGTCTCATCTGTG 639
QY 421 TTCTACAAGGAGGAGTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480
DB 640 TTCTACAAGGAGGAGTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 699
QY 481 AGCGCCATCCACTGTGCAACAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGGCCGC 540
DB 700 AGCGCCATCCACTGTGCAACAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGGCCGC 759
QY 541 AGCGCCCTGTGTCGCGGACACAACATGTGACAGCAGCAGGTTCCAGGAGTACCTGAC 600
DB 760 AGCGCCCTGTGTCGCGGACACAACATGTGACAGCAGCAGGTTCCAGGAGTACCTGAC 819
QY 601 CGCCAGGGCCGTGGCGCGTGTGGGCGAGCGTCTATCTACCCGCTCCATGAAGAAGGCCATC 660
DB 820 CGCCAGGGCCGTGGCGCGTGTGGGCGAGCGTCTATCTACCCGCTCCATGAAGAAGGCCATC 879
QY 661 GCCCAGCCCATGAAGTGGGCCAGGACCACTGGAGCCTCGCAAGAACAGCTTTGAGCTC 720
DB 880 GCCCAGCCCATGAAGTGGGCCAGGACCACTGGAGCCTCGCAAGAACAGCTTTGAGCTC 939
QY 721 TAGCGGGCTGACTTCGTCCTTGGAGGGGACTTCAGGCCCTGGCTGATCGAGATCAATTCC 780
DB 940 TAGCGGGCTGACTTCGTCCTTGGAGGGGACTTCAGGCCCTGGCTGATCGAGATCAATTCC 999
QY 781 AGCCCCACCATGACACCGCTCCAGCGGTCACCGGCCAGCTGTGTGACACAGGTTCCAGGAGTACCTG 840
DB 941 AGCCCCACCATGACACCGCTCCAGCGGTCACCGGCCAGCTGTGTGACACAGGTTCCAGGAGTACCTG 900
DB 1000 AGCCCCACCATGACACCGCTCCAGCGGTCACCGGCCAGCTGTGTGACACAGGTTCCAGGAGTACCTG 1059
QY 841 GACACCATCAGGTGGCGCGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
DB 1060 GACACCATCAGGTGGCGCGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 1119
QY 901 AGGCAG 906
DB 1120 AGCCGG 1125

RESULT 7
US-10-615-659-11
; Sequence 11, Application US/10615659
; Publication No. US2004015734A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-11

Query Match      84.1%; Score 762; DB 19; Length 3465;
Best Local Similarity 92.6%; Pred. No. 2.6e-217;
Matches 839; Conservative 0; Mismatches 0; Indels 67; Gaps 1;

QY 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGAC 60
DB 1894 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGAC 1953
QY 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAGAGAAAT 120
DB 1954 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAGAGAAAT 2013
QY 121 TACTTTTCGAGTGCAGGCTGTGCTGAATAGAAATCACTGCTGTGAACCCCTCAGACGGAC 180
DB 2014 TACTTTTCGAGTGCAGGCTGTGCTGAATAGAAATCACTGCTGTGAACCCCTCAGACGGAC 2073
QY 181 ATTGAGCGGCTCCGGAACATCTGGATTATAAGCCGCGGCAAGTCCCGGGGCCGAGAC 240
DB 2074 ATTGAGCGGCTCCGGAACATCTGGATTATAAGCCGCGGCAAGTCCCGGGGCCGAGAC 2129
QY 241 ATAGTGTGATGAGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT 300
DB 2130 ----- 2129
QY 301 TCCAGGAGCAACAAGTGGGTGTGTCAGAGAGTACATCGAGAGCGCGCTGTCTCATCTGTGAC 360
DB 2130 ---AGGAGCAACAAGTGGGTGTGTCAGAGAGTACATCGAGAGCGCGCTGTCTCATCTGTGAC 2186
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGAGTCCGGAACCCCTGAGCAATCTGG 420
DB 2187 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGAGTCCGGAACCCCTGAGCAATCTGG 2246
QY 421 TTCTACAAGGAGGAGTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480
DB 2247 TTCTACAAGGAGGAGTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 2306
QY 481 AGCGCCATCCACTGTGCAACAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGGCCGC 540
DB 2307 AGCGCCATCCACTGTGCAACAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGGCCGC 2366
QY 541 AGCCCCCTGTGTCGCGGACACAACATGTGGAACAGCAGGTTCCAGGAGTACCTGAC 600
```

Db 2367 AGCCCCCTGCTGCCGACACAAACATGTGGACAGCACAGGTTCCAGGAGTACCTGCAG 2426
QY 601 CGCCAGGGCGTGGCGCGTGTGGGCGAGCGTCATCTACCCGTCATGAAGAGGCCATC 660
Db 2427 CGCCAGGGCGTGGCGCGTGTGGGCGAGCGTCATCTACCCGTCATGAAGAGGCCATC 2486
QY 661 GCCACGCCATGAAGGTGGCCAGGACCAAGTGGAGCCCTCGAAGAACAGCTTTGAGCTC 720
Db 2487 GCCACGCCATGAAGGTGGCCAGGACCAAGTGGAGCCCTCGAAGAACAGCTTTGAGCTC 2546
QY 721 TACGGGGCTGACTTCGTCCTTGGGAGGAGCTTCAGGCCCTGGTGATCGAGATCAATTC 780
Db 2547 TACGGGGCTGACTTCGTCCTTGGGAGGAGCTTCAGGCCCTGGTGATCGAGATCAATTC 2606
QY 781 AGCCCCCACCATGACCCCGTCCAGCCGCTCACGGCCAGCTGTGTGCACAGGTGCAGGAG 840
Db 2607 AGCCCCCACCATGACCCCGTCCAGCCGCTCACGGCCAGCTGTGTGCACAGGTGCAGGAG 2666
QY 841 GACACCATCAAGGTGGCCGCTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
Db 2667 GACACCATCAAGGTGGCCGCTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 2726
QY 901 AGGCAG 906
Db 2727 AGGCAG 2732

RESULT 8
US-10-635-977-11
; Sequence 11, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2002-07-09
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-11

Query Match 84.1%; Score 762; DB 19; Length 3465;
Best Local Similarity 92.6%; Pred. No. 2.6e-217; Indels 67; Gaps 1;
Matches 839; Conservative 0; Mismatches 0
QY 1 GAGGACATCGACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGCCGAGTGGGAGGAC 60
Db 1894 GAGGACATCGACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGGCCGAGTGGGAGGAC 1953
QY 61 CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTCTCAATTCAGAAAT 120
Db 1954 CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTCTCAATTCAGAAAT 2013
QY 121 TACTTTTCCAGTGCCAGGCTCTGCTGAATAGAAATCACGCTGTGAAACCTCAGACGGAC 180
Db 2014 TACTTTTCCAGTGCCAGGCTCTGCTGAATAGAAATCACGCTGTGAAACCTCAGACGGAC 2073
QY 181 ATTGACGGGCTCGGGAACATCTGGATTATTAAGCCCGCGGCCCAAGTCCCGGGGCCGAGAC 240
Db 2074 ATTGACGGGCTCGGGAACATCTGGATTATTAAGCCCGCGGCCCAAGTCCCGGGGCCG- 2129
QY 241 ATAGTGTGATGACACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCACACACCCCTCTT 300

Db 2130 ----- 2129
QY 301 TCAGGGGACAAACAGATGGGTGGTCCAGAAAGTACATCGAGACGCCGCTGTCTCATCTGTGAC 360
Db 2130 ---AGGGACAAACAGTGGGTGGTCCAGAAAGTACATCGAGACGCCGCTGTCTCATCTGTGAC 2186
QY 361 ACCAAGTTCGACATCAGACAGATGGTTCCTGTCACGGACTGGAAACCCCTGACCATCTGG 420
Db 2187 ACCAAGTTCGACATCAGACAGATGGTTCCTGTCACGGACTGGAAACCCCTGACCATCTGG 2246
QY 421 TTCTACAGGAGAGTACTTTCGGGTTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
Db 2247 TTCTACAGGAGAGTACTTTCGGGTTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 2306
QY 481 AGCGCCCATCCACTGTGCAACCAACGCCGTCAGAAAGTACCTGAAGAAATCATGTGGGCCCG 540
Db 2307 AGCGCCATCCACTGTGCAACCAACGCCGTCAGAAAGTACCTGAAGAAATCATGTGGGCCCG 2366
QY 541 AGCCCCCTCTCTCCCGCACACAACATGTGGACACAGCACAGGTTCAGAGTACCTGCGAG 600
Db 2367 AGCCCCCTCTCTCCCGCACACAACATGTGGACACAGCACAGGTTCAGAGTACCTGCGAG 2426
QY 601 CGCCAGGGCGTGGCGCGTGTGGGGCAGCGTCATCTACCCGTCATGAAGAGGCGCATC 660
Db 2427 CGCCAGGGCGTGGCGCGTGTGGGGCAGCGTCATCTACCCGTCATGAAGAGGCGCATC 2486
QY 661 GCCCAGCCCATGAAGGTGGCCCGCAGGACCAAGTGGAGCTTACCGCTCCATGAAGAGGCGCATC 720
Db 2487 GCCCAGCCCATGAAGGTGGCCCGCAGGACCAAGTGGAGCTTCCGACAGTTCGAGCTC 2546
QY 721 TACGGGGCTGACTTCGTCCTTGGGAGGAGCTTTCAGGCCCTGTGATCGAGATCAATTC 780
Db 2547 TACGGGGCTGACTTCGTCCTTGGGAGGAGCTTTCAGGCCCTGTGATCGAGATCAATTC 2606
QY 781 AGCCCCCACCATGACCCCGTCCAGCGGTCACGGCCGAGCTGTGTCAGAGTGCAGGAG 840
Db 2607 AGCCCCCACCATGACCCCGTCCAGCGGTCACGGCCGAGCTGTGTCAGAGTGCAGGAG 2666
QY 841 GACACCATCAAGGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
Db 2667 GACACCATCAAGGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 2726
QY 901 AGGCAG 906
Db 2727 AGGCAG 2732
RESULT 9
US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-9
Query Match 83.9%; Score 760.4; DB 19; Length 1939;
Best Local Similarity 92.5%; Pred. No. 6.8e-217;
Matches 898; Conservative 0; Mismatches 1; Indels 67; Gaps 1;
QY 1 GAGGACATCGACACGTCAGCAGATGCGGTGGAGGAGCTCACTGAGGCCGAGTGGGAGGAC 60

Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGACCTCACTGAGGCCGAGTGGAGGAC 428
Qy 61 CTGACCCAGCAGTACTACTCTCTCGTTTCATGCGCATGCTTTCATCTCAATTCAGAAAT 120
Db 429 CTGACCCAGCAGTACTACTCTCTCGTTTCATGCGCATGCTTTCATCTCAATTCAGAAAT 488
Qy 121 TACTTTTCGAGTCCAGGCTCTCTGAAATAGATCACTGCTGGAACCTTCAGACCGAC 180
Db 489 TACTTTTCGAGTCCAGGCTCTCTGAAATAGATCACTGCTGGAACCTTCAGACCGAC 548
Qy 181 ATTGACGGGCTCCGGAACATCTGATTTATAAGCCCGCGGCAAGTCCCGGGCCG---- 604
Db 549 ATTGACGGGCTCCGGAACATCTGATTTATAAGCCCGCGGCAAGTCCCGGGCCG---- 604
Qy 241 ATAGTGTGATGACCGTGTGGAGAGATCTGTGAGCTGGCAGCTGCAGACCACTCTT 300
Db 605 ----- 604
Qy 301 TCAGGACACAAAGTGGTGGTCCAGAGTACATCGAGACCGCGCTCATCTGTGAC 360
Db 605 ---AGGACAAAGTGGTGGTCCAGAGTACATCGAGACCGCGCTCATCTGTGAC 661
Qy 361 ACCAAGTTCGACATCAGACAGTGGTCTCTGTCACGACTGGAAACCCCTGACCATCTGG 420
Db 662 ACCAAGTTCGACATCAGACAGTGGTCTCTGTCACGACTGGAAACCCCTGACCATCTGG 721
Qy 421 TTCTACAAGGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 480
Db 722 TTCTACAAGGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 781
Qy 481 AGCGCCATCCACTGTGTCACAAAGCCGTCAGAGTACCTGGAAGATGATGGGCGC 540
Db 782 AGCGCCATCCACTGTGTCACAAAGCCGTCAGAGTACCTGGAAGATGATGGGCGC 841
Qy 541 AGCGCCCTGTGTCGCGCACAAACATGTGACACAGCAACAGGTTCCAGGAGTACCTGCAG 600
Db 842 AGCGCCCTGTGTCGCGCACAAACATGTGACACAGCAACAGGTTCCAGGAGTACCTGCAG 901
Qy 601 CGCCAGGGCGTGGCGCGTGGGGCAGCGTCACTACCGCTCCATGAAGAGGCCATC 961
Qy 661 GCCCAGCCATGAGTGGCCAGGACCGTGGAGCTCGAGACAGCTTGGAGTCAATTC 720
Db 962 GCCCAGCCATGAGTGGCCAGGACCGTGGAGCTCGAGACAGCTTGGAGTCAATTC 1021
Qy 721 TAGCGGCTGACTTCGCTTGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTC 780
Db 1022 TAGCGGCTGACTTCGCTTGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTC 1081
Qy 781 AGCCCCACCATGCAACCGCTCCAGCCGGTCAACCGCCAGCTGTGTGACAGGTGACGAG 840
Db 1082 AGCCCCACCATGCAACCGCTCCAGCCGGTCAACCGCCAGCTGTGTGACAGGTGACGAG 1141
Qy 841 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 900
Db 1142 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 1201
Qy 901 AGGCAG 906
Db 1202 AGGCAG 1207

RESULT 10
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG542
; FILE REFERENCE: D0283A CIP

; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match 83.9%; Score 760.4; DB 19; Length 1939;
Best Local Similarity 92.5%; Pred. No. 6.8e-217;
Matches 838; Conservative 0; Mismatches 1; Indels 67; Gaps 1;

Qy 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGACCTCACTGAGGCCGAGTGGAGGAC 60
Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGACCTCACTGAGGCCGAGTGGAGGAC 428
Qy 61 CTGACCCAGCAGTACTACTCTCTCGTTTCATGCGCATGCTTTCATCTCAATTCAGAAAT 120
Db 429 CTGACCCAGCAGTACTACTCTCTCGTTTCATGCGCATGCTTTCATCTCAATTCAGAAAT 488
Qy 121 TACTTTTCGAGTCCAGGCTCTGCTGAAATAGATCACTGCTGGAACCTTCAGACCGAC 180
Db 489 TACTTTTCGAGTCCAGGCTCTGCTGAAATAGATCACTGCTGGAACCTTCAGACCGAC 548
Qy 181 ATTGACGGGCTCCGGAACATCTGATTTATAAGCCCGCGGCAAGTCCCGGGCCG---- 240
Db 549 ATTGACGGGCTCCGGAACATCTGATTTATAAGCCCGCGGCAAGTCCCGGGCCG---- 604
Qy 241 ATAGTGTGATGACCGTGTGGAGAGATCTGTGAGCTGGCAGCTGCAGACCACTCTT 300
Db 605 ----- 604
Qy 301 TCAGGACACAAAGTGGTGGTCCAGAGTACATCGAGACCGCGCTCATCTGTGAC 360
Db 605 ---AGGACAAAGTGGTGGTCCAGAGTACATCGAGACCGCGCTCATCTGTGAC 661
Qy 361 ACCAAGTTCGACATCAGACAGTGGTCTCTGTCACGACTGGAAACCCCTGACCATCTGG 420
Db 662 ACCAAGTTCGACATCAGACAGTGGTCTCTGTCACGACTGGAAACCCCTGACCATCTGG 721
Qy 421 TTCTACAAGGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 480
Db 722 TTCTACAAGGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 781
Qy 481 AGCGCCATCCACTGTGTCACAAAGCCGTCAGAGTACCTGGAAGATGATGGGCGC 540
Db 782 AGCGCCATCCACTGTGTCACAAAGCCGTCAGAGTACCTGGAAGATGATGGGCGC 841
Qy 541 AGCGCCCTGTGTCGCGCACAAACATGTGACACAGCAACAGGTTCCAGGAGTACCTGCAG 600
Db 842 AGCGCCCTGTGTCGCGCACAAACATGTGACACAGCAACAGGTTCCAGGAGTACCTGCAG 901
Qy 601 CGCCAGGGCGTGGCGCGTGGGGCAGCGTCACTACCGCTCCATGAAGAGGCCATC 660
Db 902 CGCCAGGGCGTGGCGCGTGGGGCAGCGTCACTACCGCTCCATGAAGAGGCCATC 961
Qy 661 GCCCAGCCATGAGTGGCCAGGACCGTGGAGCTCGAGACAGCTTGGAGTCAATTC 720
Db 962 GCCCAGCCATGAGTGGCCAGGACCGTGGAGCTCGAGACAGCTTGGAGTCAATTC 1021
Qy 721 TAGCGGCTGACTTCGCTTGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTC 780
Db 1022 TAGCGGCTGACTTCGCTTGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTC 1081
Qy 781 AGCCCCACCATGCAACCGCTCCAGCCGGTCAACCGCCAGCTGTGTGACAGGTGACGAG 840
Db 1082 AGCCCCACCATGCAACCGCTCCAGCCGGTCAACCGCCAGCTGTGTGACAGGTGACGAG 1141
Qy 841 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 900
Db 1142 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 1201
Qy 901 AGGCAG 906
Db 1202 AGGCAG 1207

QY 841 GACACCATCAAGTGGCCGCTGGAGCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
Db 1142 GACACCATCAAGTGGCCGCTGGAGCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 1201
QY 901 AGGCAG 906
Db 1202 AGGCAG 1207
RESULT 11
US-10-615-659-3
; Sequence 3, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-615-659-3
Query Match 80.1%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1e-206;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
Db 1 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 60
QY 241 ATAGTGTGATGACCCGCTGTGGAGAGATCTTGAGCTGGCAGCTGCAGACCAACCCTCTT 300
Db 61 ATAGTGTGATGACCCGCTGTGGAGAGATCTTGAGCTGGCAGCTGCAGACCAACCCTCTT 120
QY 301 TCCAGGGACAACAAGTGGGTGTCAGAAAGTACATCGAGACCGCGCTGCTCATCTGTGAC 360
Db 121 TCCAGGGACAACAAGTGGGTGTCAGAAAGTACATCGAGACCGCGCTGCTCATCTGTGAC 180
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGAGCTGGAACCCCTGACCATCTGG 420
Db 181 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGAGCTGGAACCCCTGACCATCTGG 240
QY 421 TTCTACAAGAGAGTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
Db 481 AGCGCCATCCACTGTGCAACAACCGCGCTCCAGAGTACCTGAAGAATGATGTGGGCCCG 540
QY 601 CGCAGGGCGGTGGCCGCTGTGGGAGCGTCACTACCCGTCATGAAGAGGCGCATC 660
Db 421 CGCAGGGCGGTGGCCGCTGTGGGAGCGTCACTACCCGCTCCATCTAAGAGAGGCGCATC 480
QY 661 GCCCAGCCATGAAGTGGCCAGGACCGTGGAGCTCGCAAGACAGCTTGA3CTC 720
Db 481 GCCCAGCCATGAAGTGGCCAGGACCGTGGAGCTCGCAAGACAGCTTGA3CTC 540

QY 721 TACGGGGCTGACTTCGTCTCTGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTCC 780
Db 541 TACGGGGCTGACTTCGTCTCTGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTCC 600
QY 781 AGCCCCACCATGACACCCGTCACGCGGTCACGGCCACGCTGTGTCACAGGTGCAGGAG 840
Db 601 AGCCCCACCATGACACCCGTCACGCGGTCACGCGCCACGCTGTGTCACAGGTGCAGGAG 660
QY 841 GACACCATCAAGTGGCCGCTGGAGCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
Db 661 GACACCATCAAGTGGCCGCTGGAGCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 720
QY 901 AGGCAG 906
Db 721 AGGCAG 726
RESULT 12
US-10-635-977-3
; Sequence 3, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-635-977-3
Query Match 80.1%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1e-206;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
Db 1 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 60
QY 241 ATAGTGTGATGACCCGCTGTGGAGAGATCTTGAGCTGGCAGCTGCAGACCAACCCTCTT 300
Db 61 ATAGTGTGATGACCCGCTGTGGAGAGATCTTGAGCTGGCAGCTGCAGACCAACCCTCTT 120
QY 301 TCCAGGGACAACAAGTGGGTGTCAGAAAGTACATCGAGACCGCGCTGCTCATCTGTGAC 360
Db 121 TCCAGGGACAACAAGTGGGTGTCAGAAAGTACATCGAGACCGCGCTGCTCATCTGTGAC 180
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGAGCTGGAACCCCTGACCATCTGG 420
Db 181 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGAGCTGGAACCCCTGACCATCTGG 240
QY 421 TTCTACAAGAGAGTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
Db 241 TTCTACAAGAGAGTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 300
QY 481 AGCGCCATCCACTGTGCAACAACCGCGCTCCAGAGTACCTGAAGAATGATGTGGGCCCG 540
Db 301 AGCGCCATCCACTGTGCAACAACCGCGCTCCAGAGTACCTGAAGAATGATGTGGGCCCG 360
QY 541 AGCGCCCTGCTGCCCGCACACAACATGTGGACACAGCAAGTTCAGGAGTACCTGTCAG 600

Db 361 AGCCCCCTGCTGCCCGCACACATGTGACACGACAGAGTTCCAGGAGTACCTGCAG 420
Qy 601 CGCCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCAATCTACCGTCCATGAAGAAGGCCATC 660
Db 421 CGCCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCAATCTACCGTCCATGAAGAAGGCCATC 480
Qy 661 GCCCAGCCCATGAGGTGGCCAGGACCAAGTGGAGCTCGCAAGACAGCTTTGAGCTC 720
Db 481 GCCCAGCCCATGAGGTGGCCAGGACCAAGTGGAGCTCGCAAGACAGCTTTGAGCTC 540
Qy 721 TAGGGGCTGACATTCGCTTCCTTGGAGGAGCACTTCAAGGCCCTGGGTGATCGAGATCAATTC 780
Db 541 TAGGGGCTGACATTCGCTTCCTTGGAGGAGCACTTCAAGGCCCTGGGTGATCGAGATCAATTC 600
Qy 781 AGCCCCACCATGACCCCGTCCAGCCCGTCAAGGCCAGCTGTGTGACAGGTGCAAGGAG 840
Db 601 AGCCCCACCATGACCCCGTCCAGCCCGTCAAGGCCAGCTGTGTGACAGGTGCAAGGAG 660
Qy 841 GACACCATGAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
Db 661 GACACCATGAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 720
Qy 901 AGGCAG 906
Db 721 AGGCAG 726
RESULT 13
US-10-029-386-24894/c
; Sequence 24894, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24894
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022327.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.35
; OTHER INFORMATION: SWISSPROT HIT: P38160, EVALUE 5.00e-04
; OTHER INFORMATION: EST HUMAN HIT: BE82753.1, EVALUE 3.00e-06
; OTHER INFORMATION: NT HIT: g114726145, EVALUE 8.00e-09
US-10-029-386-24894
Query Match 47.1%; Score 427; DB 16; Length 490;
Best Local Similarity 100.0%; Pred. No. 2.8e-117;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 480 CAGCGCCATCCACTGTGCAACACCGCGTCCAGAGTACCTGAAGAATGATGTGGCCG 539
Db 427 CAGCGCCATCCACTGTGCAACACCGCGTCCAGAGTACCTGAAGAATGATGTGGCCG 368
Qy 540 CAGCCCCCTGCTGCCCGCACACAATGTGGACAGCAGCAGGTTCAGGAGTACCTGCA 599
Db 367 CAGCCCCCTGCTGCCCGCACACAATGTGGACAGCAGCAGGTTCAGGAGTACCTGCA 308
Qy 600 GCGCCAGGGCCGTGGCGCGTGTGGGCGACGCTCATCTACCGTCCATGAAGAAGGCCAT 659
Db 307 GCGCCAGGGCCGTGGCGCGTGTGGGCGACGCTCATCTACCGTCCATGAAGAAGGCCAT 248
Qy 660 CGCCACGCATGAGGTGGCCCGCAGGACCAAGTGGAGCTTCGCAAGACAGCTTTGAGCT 719

Db 247 CGCCACGCCATGAGGTGGCCCGCAGGACCAAGTGGAGCTTCGCAAGACAGCTTTGAGCT 188
Qy 720 CTACGGGGCTGACTTCGTCCTTGGGAGGAGACTTTCAGGCCCTTGCTGATCGAGATCAATTC 779
Db 187 CTACGGGGCTGACTTCGTCCTTGGGAGGAGACTTTCAGGCCCTTGCTGATCGAGATCAATTC 128
Qy 780 CAGCCCCACCATGACCCCGTCCAGCGCGGTTCACGGGCCAGCTGTGTGCAAGGTGCAAGG 839
Db 127 CAGCCCCACCATGACCCCGTCCAGCGCGGTTCACGGGCCAGCTGTGTGCAAGGTGCAAGG 68
Qy 840 GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTG 899
Db 67 GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTG 8
Qy 900 GAGGCAG 906
Db 7 GAGGCAG 1
RESULT 14
US-10-723-860-631/c
; Sequence 631, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 631
; LENGTH: 101270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-631
Query Match 47.1%; Score 427; DB 20; Length 101270;
Best Local Similarity 100.0%; Pred. No. 1e-116;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 480 CAGCGCCATCCACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGAATGATGTGGCCG 539
Db 44821 CAGCGCCATCCACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGAATGATGTGGCCG 44762
Qy 540 CAGCCCCCTGCTGCCCGCACACAATGTGGACCAAGGTTCAGGAGTACCTGCA 599
Db 44761 CAGCCCCCTGCTGCCCGCACACAATGTGGACCAAGGTTCAGGAGTACCTGCA 44702
Qy 600 GCGCCAGGGCCGTGGCGCGTGTGGGCGAGCGTCACTACCGTCCATGAGAGGCCAT 659
Db 44701 GCGCCAGGGCCGTGGCGCGTGTGGGCGAGCGTCACTACCGTCCATGAGAGGCCAT 44642
Qy 660 CGCCACGCATGAGGTGGCGCGCAGGACCAAGTGGAGCTTCGCAAGAAACAGCTTTGAGCT 719
Db 44641 CGCCACGCATGAGGTGGCGCGCAGGACCAAGTGGAGCTTCGCAAGAAACAGCTTTGAGCT 44582
Qy 720 CTACGGGGCTGACTTCGTCCTTGGGAGGAGCTTCAGGCCCTTGCTGATCGAGATCAATTC 779
Db 44581 CTACGGGGCTGACTTCGTCCTTGGGAGGAGCTTCAGGCCCTTGCTGATCGAGATCAATTC 44522
Qy 780 CAGCCCCACCATGACCCCGTCCAGCGCGGTTCACGGGCCAGCTGTGTGCAAGGTGCAAGG 839
Db 44521 CAGCCCCACCATGACCCCGTCCAGCGCGGTTCACGGGCCAGCTGTGTGCAAGGTGCAAGG 44462
Qy 840 GGACACCATCAAGGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTG 899
Db 44461 GGACACCATCAAGGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTG 44402

QY 900 GAGGACAG 906
Db 44401 GAGGACAG 44395

RESULT 15
US-10-275-595A-39
; Sequence 39, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, DYUNG Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 2380
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CB1
US-10-275-595A-39

Search completed: September 24, 2005, 16:48:43
Job time : 701.935 secs

Query Match 44.0%; Score 398.8; DB 18; Length 2380;
Best Local Similarity 66.3%; Pred. No. 1.1e-108;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 3 GGACATCGACAGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGGAGACCT 62
Db 473 GGACATCGACAGGACCTGGAGGCCCGCTGTACTCACCCTGGAGGCGTGTCCCTCTT 532

QY 63 GACCAGCAGTACTACTCCCTGTTTCATGCGATGCTTTCATCTCCAATTCAGAAATTA 122
Db 533 CCTCAGCCGCTACTACCAAGTGGTCCACGAAGGGCAGAACTCAGGCACCTCGACACTCA 592

QY 123 CTTTTCGCAAGTGCAGGCTCTGCTGAATAGAAATCAGCTCTGTGAACCTTCAGACGGACAT 182
Db 593 GGTCCAGCGCTGTGAGGACATCTGCAGCAGCTGCAGGCCGTGGTATCCCCAGATAGACAT 652

QY 183 TGACGGGCTCCGGAACATCTGGATATTAAGCCGGGCAAGTCCCGGGCCGAGACAT 242
Db 653 GGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAGTCCCGTGGACGAGGCAT 712

QY 243 AGTGTGCATGGACCGGTGTGGAGAGATCCTGGAGCTGGCAGCTGCAGACCCCTCTTTC 302
Db 713 CATGTGCATGGACCCACCTCGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGTGAT 772

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	398.8	44.0	3001	4	US-09-620-312D-494	Sequence 494, App
2	52.6	5.8	1272	4	US-09-252-991A-471	Sequence 2471, Ap
3	52.6	5.8	1449	4	US-09-252-991A-3200	Sequence 2200, Ap
4	52.6	5.8	2790	4	US-09-252-991A-2632	Sequence 2632, Ap
5	48.6	5.4	2277	1	US-08-676-967-5	Sequence 5, Appli
6	48.6	5.4	2277	1	US-08-676-974-5	Sequence 5, Appli
7	48.6	5.4	2277	2	US-09-098-487-5	Sequence 5, Appli
8	48.2	5.3	2004	1	US-08-471-033-18	Sequence 18, Appl
9	48.2	5.3	2004	2	US-08-471-044-18	Sequence 18, Appl
10	48.2	5.3	2004	2	US-08-463-483A-18	Sequence 18, Appl
11	48.2	5.3	2004	2	US-08-471-046A-18	Sequence 18, Appl
12	48.2	5.3	2004	2	US-08-470-566B-18	Sequence 18, Appl
13	48.2	5.3	2004	2	US-08-469-334-18	Sequence 18, Appl
14	48.2	5.3	2004	3	US-09-300-529-18	Sequence 18, Appl
15	48.2	5.3	2655	1	US-08-471-033-17	Sequence 17, Appl
16	48.2	5.3	2655	2	US-08-471-044-17	Sequence 17, Appl
17	48.2	5.3	2655	2	US-08-463-483A-17	Sequence 17, Appl
18	48.2	5.3	2655	2	US-08-471-046A-17	Sequence 17, Appl
19	48.2	5.3	2655	2	US-08-470-566B-17	Sequence 17, Appl
20	48.2	5.3	2655	2	US-08-469-334-17	Sequence 17, Appl
21	48.2	5.3	2655	3	US-09-300-529-17	Sequence 17, Appl
22	48	5.3	5314	4	US-09-620-312D-138	Sequence 138, App
23	47.6	5.3	4884	4	US-09-252-991A-12126	Sequence 12126, A
24	47.6	5.3	4884	4	US-09-252-991A-12292	Sequence 12292, A
25	45.8	5.1	1515	4	US-09-252-991A-9184	Sequence 9184, Ap
26	45.8	5.1	9556	4	US-09-902-540-982	Sequence 982, App
27	45.6	5.0	1047	4	US-09-252-991A-14678	Sequence 14678, A

Db

US-09-620-312D-494

Db

```
QY 63 GACCCAGCAGTACTACTCCTCCTGTTATGAGCGGATGCTTTTATCTCCTCAATTCAGAAATTA 122
    |||||
Db 1423 CTTCCAGCGCTACTACCAAGTGGTCCAGAGGGGAGAACTCAGGCACCTCGACACTCA 1482

QY 123 CTTTTCGAGTCCAGGCTCTGCTGATAGATACAGTCTGTGAACTCCTCAGACGACAT 182
    |||||
Db 1483 GGTCCAGCGCTGTGAGGACATCTCGACGAGCTGCGAGGCGGTGATACCCCGAGATAGACAT 1542

QY 183 TGACGGGCTCCGGAACATCTGGATTTAAAGCCCGGCGCAAGTCCCGGGCGGCGAGACAT 242
    |||||
Db 1543 GGAAGGGATCGCAACATCTGGATCTGGAAGCGAGGACCAAGTCCCGGGAGGAGCAT 1602

QY 243 AGTGTGCATGAGACCGTGTGAGAGAGATCTCGAGCTGGAGCTGCAGACACACCTCTTTC 302
    |||||
Db 1603 CATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACGGCAAACCCGTTGTGAT 1662

QY 303 CAGGGACACAGTGGTGGTCCAGAGTACATCGAGACGCCGCTCTCATCTGTGACAC 362
    |||||
Db 1663 GAAGGACCGCAAGTGGTGGTGCAGAGATATATTGAGCGGCGCCCTCTCATCTTTGGCAC 1722

QY 363 CAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGGAAACCCCTCGACCATCTGGTT 422
    |||||
Db 1723 CAAGTTTGACCTCAGACAGTGGTTCCTGTTAACTGACTGGAACCACTTACCGTGTGGTT 1782

QY 423 CTAAAGAGAGTACTTCCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAG 482
    |||||
Db 1783 CTACCGCAGACGCTATATCCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACTGGACAA 1842

QY 483 CGCCATCCACCTGTGCAACAAACGCGCTCCAGAGTACCTGAAGATGATGTGGGCGCGAG 542
    |||||
Db 1843 CTCAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCCATCGGCA 1902

QY 543 CCCCTGCTGCCGCGCACAAACATGTGGACCAAGTTCAGGAGTCTCCAGGAGTACTCTGCAGG 602
    |||||
Db 1903 TCCACTGCTTCCGCGACAGACATGTGCTGTAGCCAGAGTTCAGGCCCACTCGACGA 1962

QY 603 CCAGGCGGTGGCGCGTGTGGGCGAGCTCATCTACCCGTCATGGAAGGCCATCGC 662
    |||||
Db 1963 GATGGGTGCCCAATGCTTGGTCCACCATCATCTGTCGCTGGCATGAAGGATGCTGTGAT 2022

QY 663 CCAGCCATGAAGTGGCGCCAGGACACATGAGCCCTCCAGACAGCTTTGAGCTCTA 722
    |||||
Db 2023 CCAGGCACATCGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTA 2082

QY 723 CGGGGTGACTTCTGCTTGGGAGGAGTTCAGGCCCTCGGCTGATCGAGATCAATTCCAG 782
    |||||
Db 2083 TGGCGTGACTTCTGCTTGGGAGGAGTTCAGGCCCTCGGCTGATGATGATCAACGCCAG 2142

QY 783 CCCCACCATGCAACCGTCCACGCCGCTCAGGCCCAAGCTGTGTGCAAGGTGCAAGAGA 842
    |||||
Db 2143 CCCCACGATGGCAACCTCCACAGCAGTCACTGCGCGCTCTGTGCTGGCGTGCAAGCTGA 2202

QY 843 CACCATCAAGTGGCGGTGGACCGCA 868
    |||||
Db 2203 CACCCTGCGGTGGTCAATTGACCGGA 2228
```

```
RESULT 2
US-09-252-991A-2471
; Sequence 2471, Application US/09252991A
; Patent No. 6551795
; ORGANISM: Pseudomonas aeruginosa
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```

```
; SEQ ID NO 2471
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2471

Query Match      5.8%; Score 52.6; DB 4; Length 1272;
Best Local Similarity 48.5%; Pred. No. 0.00059;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 177 GCACATTGACGGGCTCCGGAACATCTGGATTTAAAGCCCGGCGCAAGTCCCGGGCGG 236
    |||||
Db 756 GGAATATGCTCGCGACCGCCAGGTACGGACGATTTCCGGCACCGGCGAGGACGATCTT 815

QY 237 AGACATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGGAGCTGCAGACACC 296
    |||||
Db 816 CGACGTACGGATGGTTCGACCGGCGGCGGTGGAGGTGGTTCGAGTAGGCCAGGC 875

QY 297 TCTTTTCCAGGGACAACAAGTGGTGGTCCAGAAATGATCATCGAGACGCCGCTCATCTG 356
    |||||
Db 876 GGTGACGCTGGAAGTTCGAGTTCGAGTTCGCCAGGACATCGAGCGGCTGATCTCGGCTT 935

QY 357 TGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGGAACCCCTGACCAT 416
    |||||
Db 936 CATGATCAAGGACCGCCTCGGCCAGCGATGTACGGGATCAATACCCACCGCTGGACAA 995

QY 417 CTGGTTCTACAGGAGAGTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTCGACAAGC 475
    |||||
Db 996 GSGCGTACCAGACCTGAAAGCGGCGGAGGATCATCTACCGTTGCGCTTCGACATGC 1054

RESULT 3
US-09-252-991A-2200/c
; Sequence 2200, Application US/09252991A
; Patent No. 6551795
; ORGANISM: Pseudomonas aeruginosa
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2200
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2200
```

```
Query Match      5.8%; Score 52.6; DB 4; Length 1449;
Best Local Similarity 48.5%; Pred. No. 0.00059;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 177 GCACATTGACGGGCTCCGGAACATCTGGATTTAAAGCCCGGCGCAAGTCCCGGGCGG 236
    |||||
Db 562 GGAATATGCTCGCGACCGCCAGGTACGGACGATTTCCGGCACCGGCGAGGACGATCTT 503

QY 237 AGACATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGGAGCTGCAGACACC 296
    |||||
Db 502 CGACGTACGGATGGTTCGACCGGCGGCGGTGGAGGTGGTTCGAGTAGGCCAGGC 443

QY 297 TCTTTTCCAGGGACAACAAGTGGTGGTCCAGAAATGATCATCGAGACGCCGCTCATCTG 356
    |||||
Db 442 GGTGACGCTGGAAGTTCGAGTTCGAGTTCGCCAGGACATCGAGCGGCTGATCTCGGCTT 383

QY 357 TGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGGAACCCCTGACCAT 416
    |||||
Db 382 CATGATCAAGGACCGCCTCGGCCAGCGGATGTACGGGATCAATACCCACCGCTGGACAA 323
```

QY 417 CTGTTTACAGGAGAGTTACTTGGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 475
Db 322 GCGCTACCGACTGAAGCGCGGAGCGGATCACCTACCGCTTCGCTTCGACATGC 264

RESULT 4

US-09-252-991A-2632
; Sequence 2632, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2632
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2632

Query Match 5.8%; Score 52.6; DB 4; Length 2790;
Best Local Similarity 48.5%; Pred. No. 0.00076;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 177 GGACATTGACGGGCTCCGGACATCTCGATTATTAAGCCCGCCCAAGTCCCGGGCGG 236
Db 895 GGAATGCTCGCGACGCCAGGTTACGGACGATTTCCGGACACCGCGGAGCGGATCCT 954
QY 237 AGACATAGTGTGACATGACCGTGTGGAGGAGATCCTGGAGTGCAGCTGCAGACACACC 296
Db 955 CGACGTACGATGTGCACCGCGGAGCGGCGCTGGAGTGTGAGTAGGCGGCGG 1014
QY 297 TCTTTCAGGGACAAAGTGGTGGTCCAGAAAGTACATCGAGACGCGCTGTCTCATCTG 356
Db 1015 GGTGACGCTGGAAGTCGAGGTCGAGGTTGCGCAGGACATCGAGCGGCTGCTCGGCTT 1074
QY 357 TGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCAGGACGGAACCCCTGACAT 416
Db 1075 CATGATCAAGACCGCTCGCGCCAGCGGATGTACGGGATCAATACCCACCGCTGGACAA 1134
QY 417 CTGTTTCTACAGGAGAGTTACTTGGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 475
Db 1135 GCGCTACCGACTGAAGCGCGGAGCGGATCACCTACCGCTTCGCTTCGACATGC 1193

RESULT 5

US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-5

Query Match 5.4%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.0079;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 465 CCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAGCGCGTCCAGAACTACCTGAA 524
Db 1437 CGAATCATCTTCTGTGAGCGCGACCGGCTGTGCTGTCACAACTGCCAAGGCGGTTGA 1496
QY 525 GAATGATGTGGCGCGGACGCCCTGTGTCGCCGACACAAATGTGGACACGACACAGGTT 584
Db 1497 CGACAAGCAGCTGCGCAAGCTGTGCTGAGCGGCCACCGAGCGGAGAGGCGGTCGCAT 1556
QY 585 CAGGAGTACTGACAGCGCGGCGGCGGCTGTGGGCGAGCGTCTATCTACCGGTC 644
Db 1557 CAAGGAGTGCCTGATGCGGACCTGAAGGCGGTGACCGCAACATGAAGGCGCAGAG 1616
QY 645 CATGAAGAAGGCGCATCGCCACCGCATGAAGTGGCCGAGGACACCGTGGAGGCTCGCAA 704
Db 1617 CTTGGGCTAGCCTTTCGCGAGTTCAGGAGCAGCAGCAGCGCTGAGGCGGCTCGGCT 1676
QY 705 GAACAGC 711
Db 1677 GATCAAC 1683

RESULT 6

US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-098-487-5

Query Match 5.4%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.0079;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 465 CTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAGCGCGTCCAGAAGTACCTGAA 524
DB 1437 CCAGAACATCTTCGTGAGCGCCACCGCCCTGTGCTGCAACCTGCCAAGCGCGTGA 1496

QY 525 GAATGATGTGGCGCGCAGCGCCCTGTGCTGCCCGCACACAAACATGTGGACACCAACAGGTT 584
DB 1497 CGACAAAGCAGTGCAGCAAGCTGTCTGAGCGCCACCGCGGAGAGAGCGGTGCGCAT 1556

QY 585 CAGGAGTACCTGCAGCGCCAGCGCGCGTGTGGCGCAGCGTGTATCTACCCGTC 644
DB 1557 CAGGAGTGCAGCGCGTGTATCTGAGCGCCACCGCGGAGAGAGCGGTGCGCAT 1616

QY 645 CATGAAGAAGCGCCATCGCCACGCCATGAAGTGGCGCCAGGACCACTGAGCGCTCGCAA 704
DB 1617 CTGGGCTAGCGCTTCGCCAGTTCAGGAGCAGGACGAGCGCCCTGAAGCGCGTGCCT 1676

QY 705 GAACAGC 711
DB 1677 GATCAAC 1683

RESULT 7
US-09-098-487-5
Sequence 5, Application US/09098487
Patent No. 5917025
GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: cDNA
US-09-098-487-5

Query Match 5.4%; Score 48.6; DB 2; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.0079;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 465 CTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAGCGCGTCCAGAAGTACCTGAA 524
DB 1437 CCAGAACATCTTCGTGAGCGCCACCGCCCTGTGCTGCAACCTGCCAAGCGCGTGA 1496

QY 525 GAATGATGTGGCGCGCAGCGCCCTGTGCTGCCCGCACACAAACATGTGGACACCAACAGGTT 584
DB 1497 CGACAAAGCAGTGCAGCAAGCTGTCTGAGCGCCACCGCGGAGAGAGCGGTGCGCAT 1556

QY 585 CAGGAGTACCTGCAGCGCCAGCGCGCGTGTGGCGCAGCGTGTATCTACCCGTC 644
DB 1557 CAGGAGTGCAGCGCGTGTATCTGAGCGCCACCGCGGAGAGAGCGGTGCGCAT 1616

QY 645 CATGAAGAAGCGCCATCGCCACGCCATGAAGTGGCGCCAGGACCACTGAGCGCTCGCAA 704
DB 1617 CTGGGCTAGCGCTTCGCCAGTTCAGGAGCAGGACGAGCGCCCTGAAGCGCGTGCCT 1676

QY 705 GAACAGC 711
DB 1677 GATCAAC 1683

RESULT 8
US-09-471-033-18
Sequence 18, Application US/08471033
Patent No. 5770696
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-033-18

Query Match 5.3%; Score 48.2; DB 1; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 291 CCACCTCTTTCCAGGACCAACAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCT 350
DB 720 CCACCCCATCACCTGACAGAGAGAGTGGAGAACCTCTGACACCAAGCCCATGAT 779
QY 351 CATCTGTGACCAAGTTGACATCATGACAGTGGTCTCTGTCACGAGCTGGAACCCCT 410
DB 780 GCTGGAGACCAACAGACCGCGCTCTACAAGATCAAGGACACCCACGCGCAACATCGT 839
QY 411 GACCATCTGTTTACAGAGAGAGTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 470
DB 840 GACCGCGCGAGTGGAAACCGCGTGTATCCAGCAGATCAAGGCCAAGACCCCGCAGCATCAT 899
QY 471 CAAGCTGGACAGCGCATCCACCTGTGCAACAACGCGCTCCAGAGTACCTGAAGATGA 530
DB 900 CGTCGACGACGCGCGCTGTGCGGAGAGCGGTGGCGCCGAGACTTACGAGAACCC 959
QY 531 TGTGGCGCGAGCCCTGCTGCGCGCACACACATGTGGACGACGACCCAGGTTCCAGGA 590
DB 960 CGAGGACCAACCCCGACCTGACCTTGAAGGACGCGCTGAAGTGAGCTACCCGACGA 1019
QY 591 GTACTGACGCGCAGCGCGCTGCGCGCGTGTGGGAGAGGCTATCTACCCGTCATGA 650
DB 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACCAAGCCCATCTACGAGACGCGT 1079
QY 651 GAAGGCGCATCGCCAGCCATGAAGTGGCGCCAGGACCACTGGAGCTCGCAAGAACAG 710
DB 1080 GATGACCTATCTAGACGAGAACACCGCCAGAGAGGTGACCAAGCAGCTGAACGACCCAC 1139
QY 711 CTTTGTAGCTTACGGGCTGACTTCGTCCTTGGGAGGAGCTTACGGCCCTGGCTGATCGA 770
DB 1140 CGCAAGTTCAAGGACGTGAGCCACCTGTACTAGACGTGAAGCTGACCCCGCAAGATGA 1199
QY 771 GATCA 775
DB 1200 GACCA 1204

RESULT 9

US-08-471-044-18
Sequence 18, Application US/08471044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kozichka, N. Kristy
APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-044-18

Query Match 5.3%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 291 CCACCTCTTTCCAGGACCAACAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCT 350
DB 720 CCACCCCATCACCTGACAGAGAGAGTGGAGAACCTCTGACACCAAGCCCATGAT 779
QY 351 CATCTGTGACCAAGTTGACATCATGACAGTGGTCTCTGTCACGAGCTGGAACCCCT 410
DB 780 GCTGGAGACCAACAGACCGCGCTCTACAAGATCAAGGACACCCACGCGCAACATCGT 839
QY 411 GACCATCTGTTTACAGAGAGAGTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 470
DB 840 GACCGCGCGAGTGGAAACCGCGTGTATCCAGCAGATCAAGGCCAAGACCCCGCAGCATCAT 899
QY 471 CAAGCTGGACAGCGCATCCACCTGTGCAACAACGCGCTCCAGAGTACCTGAAGATGA 530
DB 900 CGTCGACGACGCGCGCTGTGCGGAGAGCGGTGGCGCCGAGACTTACGAGAACCC 959

Qy	531	TGTTGGCGGAGCCCTGTGTCGGGCACACAATGTGGACCAGCACCAAGTTCAGGA	590
Db	960	CGAGGACAAAGACCCCAGCTTGACCTTGAAGGACGCCCTGAAGGTGACTACCCCGACGA	1019
Qy	591	GTACTTCAGCGCAGGCGCGTGTGGGGCAGCGTCATCTACCCGCTCCATGAA	650
Db	1020	GATCAAGGAGATCGAGGGCTGTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT	1079
Qy	651	GAAGGCCATCGCCCCAGCCCATGAAGTGGCCAGGACACGTGGAGCGCTCGCAAGAACAG	710
Db	1080	GATGACCTATCTAGACGAGAACCCGCCAAGGAGTGACCAAGCAGTGAACGACCCAC	1139
Qy	711	CTTTGAGCTCTACGGGGCTGACTTCGTCTTTGGAGGGAATTACAGGCCCTGGGTGATCGA	770
Db	1140	CGGCAAGTTCAAGGACGTGAGCCACCTGTGACGAGCTGAAGCTGACCCCCCAAGATGAACGT	1199
Qy	771	GATCA	775
Db	1200	GACCA	1204

```

RESULT 10
US-08-463-483A-18
; Sequence 18, Application US/08463483A
; Patent No. 5849870
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50

```

Qy 591 GTACTCTGACGGCGCAGGCGCGTGGCGCCGTGTGGGGGAGCGTCATCTACCCGTCCATGA 650
Db 1020 GATCAAGGAGATCGAGGGCGCTGCTGTACTACAAGAACAAGCCCCATCTACGAGAGCAGCGT 1079
Qy 651 GAAGGCCATCGCCACGCAGCATGAAGTTGGCCCAAGGACCACGTGGAGCCTCGCAAGACAG 710
Db 1080 GATGACCTATTCTAGACGAGAACACCGCCAAAGGAGGTGACCAAGCAGCTGAACACACCAC 1139
Qy 711 CTTTGAGCTCTACGGGGCGTACATTCGTCTTGGGAGGAGCTTCAGGCCCTGGGTGATCGA 770
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAGCTGACCCCACCAAGATGAACGT 1199
Qy 771 GATCA 775
Db 1200 GACCA 1204

RESULT 12
US-08-470-566B-18
; Sequence 18, Application US/08470566B
; Patent No. 5872212
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5872212artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,566B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

1 COUNTRY: USA
2 ZIP: 10532
3 COMPUTER READABLE FORM:
4 MEDIUM TYPE: Floppy disk
5 COMPUTER: IBM PC compatible
6 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: PatentIn Release #1.0, Version #1.30B
8 CURRENT APPLICATION DATA:
9 FILING DATE: 06-JUN-1995
10 APPLICATION NUMBER: US/08/469,334
11 CLASSIFICATION: 800
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: 08/463,483
14 FILING DATE:
15 APPLICATION NUMBER: US 08/314,594
16 FILING DATE: 09-SEP-1994
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: US 08/218,018
19 FILING DATE: 23-MAR-1994
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 08/037,057
22 FILING DATE: 25-MAR-1993
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Spruill, W. Murray
25 REGISTRATION NUMBER: 32,943
26 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: 919-541-8615
29 TELEFAX: 919-541-8689
30 INFORMATION FOR SEQ ID NO: 18:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 2004 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36 MOLECULE TYPE: DNA (genomic)
37 HYPOTHETICAL: NO
38 ANTI-SENSE: NO
39 FEATURE:
40 NAME/KEY: misc.feature
41 LOCATION: 1..2004
42 OTHER INFORMATION: /note= "Maize optimized DNA
43 OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
44 US-08-469-334-18
45
46 Query Match 5.3%; Score 48.2; DB 2; Length 2004;
47 Best Local Similarity 43.7%; Pred. No. 0.0096;
48 Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

```

[illegible]

QY 651 GAAGGCGATCGCCACGCGATGAGGTGGCCAGGACACGCTGGAGCGCTCGCAAGAACAG 710
|||
Db 1080 GATGACCTATCTAGACGAGAACACCGCCAGGAGGTGACCAAGCAGCTGAACGACACAC 1139
|||
QY 711 CTTTGAGCTCTACGGGGCTGACTTCGTCCTTTGGAGGAGCTTCAGGGCCCTGGCTGATCGA 770
|||
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
|||
QY 771 GATCA 775
|||
Db 1200 GACCA 1204
|||

RESULT 14

US-09-300-529-18
; Sequence 18, Application US/09300529
; Patent No. 6066783
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6066783artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,529
; FILING DATE: TBA
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,334
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19506L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
; US-09-300-529-18

Query Match 5.3%; Score 48.2; DB 3; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 291 CCACCCCTCTTCCAGGAGCAACAAGTGGGTGGTTCAGAAAGTACATCCAGAGCCGCGCTGCT 350
|||
Db 720 CCACCCCATCACCTCGAACAAGAGCAGAGGTGGACAACTCTGTGAACAACAAGCCCATGAT 779
|||
QY 351 CATCTGTGACACCAAGTTCCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAAACCCCT 410
|||
Db 780 GCTGGAGACCAACAGACCCGCGGTCTTCAAGATCAAGGACACCCACCGCAACATCGT 839
|||
QY 411 GACCATCTGTTCTTCAAGAGGAGATTACTTGCCTGTTCTCAACTCAGCGCTTCTCCCTGGA 470
|||
Db 840 GACCGGCGGAGTGGACCGCGGTGATCCAGCAGATCAAGGCCAAGACCCGCGCATCAT 899
|||
QY 471 CAAGCTGGACAGCGCCATTCACCTGTGTGCAACAACCGCGTCCAGAAGTACCTGGAAGAATGA 530
|||
Db 900 CGTCGACGACGGCGAGCGCGTGGCCGAGAGCGCGTGGCCGCAAGGACTACGAGAACC 959
|||
QY 531 TGTGGGCGGACGCCCTGCTGCGCGCACACAACATGTGGACCAAGCAGCAGGTTCCAGGA 590
|||
Db 960 CGAGGACAAGACCCCGCTGACCTTGAAGGACGCCCTGAAGCTGAGCTTACCCCGAGCA 1019
|||
QY 591 GTACCTGCAGCGCAGGGCGTGGCGCGTGTGGGGCAGCGTTCATCTACCCGTCATGAA 650
|||
Db 1020 GATCAAGGAGATCGAGGCGCTGCTGTACTACAAGAACAAGCCCATCTACGAGAGCAGCGT 1079
|||
QY 651 GAAGGCGATCGCCACGCGCATGAAGTGGGCCAGGACACGCTGGAGCCTCGCAAGAACAG 710
|||
Db 1080 GATGACCTATCTAGACGAGAACACCCGCAAGGAGGTGACCAAGCAGCTGAACGACAC 1139
|||
QY 711 CTTTGAGCTCTACGGGCTGACTTCGTCCTTTGGAGGAGCTTCAGGCGCTGGCTGATCGA 770
|||
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
|||
QY 771 GATCA 775
|||
Db 1200 GACCA 1204
|||

RESULT 15

US-08-471-033-17
; Sequence 17, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA

ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..2652
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for 100 kd VIP1A(a) protein from AB78"
US-08-471-033-17

Query Match 5.3%; Score 48.2; DB 1; Length 2655;
Best Local Similarity 43.7%; Pred. No. 0.011;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 291 CCACCCCTCTTCCAGGCAACAAGTGGGTGCCAGAGTACATCGAGACGCGCTGCT 350
DB 1371 CCACCCCATCACCTTGAACAAGAGCAGGTGGACAACCTGTGTGAACAACAGCCCATGAT 1430
QY 351 CATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCTCTCGTCACGGAATGGAACCCCT 410
DB 1431 GCTGGAGACCAACGACGCGGCTCTACAAGATCAAGGACACCCACGGCAACATCGT 1490
QY 411 GACCATCTGGTCTACAAGAGAGTTACTTGGGTTCTCAATCAGCGCTTCTCCCTGGA 470
DB 1491 GACCGCGCGAGTGAACGGCTGTATCCAGCAGATCAAGGGCAAGACCGCCAGCATCAT 1550
QY 471 CAAGCTGGACAGCGCCATCCACTGTGCAACAACGCGCTCCAGAGTACTCTGAAGATGA 530
DB 1551 CGTCGACGACGGGAGCGGTGGCCGAGAGCGGTGGCGCCGAGGACTACGAGAACCC 1610
QY 531 TGTGGGCGCAGCCCTGTGTCGCCCAACAACATGTGGAACGACGACCCAGGTTCAGGA 590
DB 1611 CGAGGACAAGACCCCGAGCTGACCTGAAGGACGCCCTGAAGCTGAGTACCCGACGA 1670
QY 591 GTACTGTGACGCGCAGGCGCGTGGCGCGTGTGGGGCAGCGTCACTACCGTCCATGAA 650
DB 1671 GATCAAGGAGATCGAGGGCTGTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT 1730
QY 651 GAAGGCCATCGCCCGCATGAAGTGGCCCGAGGACACGAGTGGAGCCTTCGGAAGAACAG 710
DB 1731 GATGACCTATCTAGACGAGAACACCCGCCAAGGAGGTGACCAAGCAGCTGAACGACACCAC 1790

QY 711 CTTTGAGCTCTACGGGGCTGACTTCGTCTGGGAGGAGACTTTCAGGCCCTGGCTGATCGA 770
DB 1791 CGGCAAGTTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1850
QY 771 GATCA 775
DB 1851 GACCA 1855

Search completed: September 24, 2005, 11:30:36
Job time : 166.553 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:08:28 ; Search time 3481.78 Seconds
(without alignments)
9904.766 Million cell updates/sec

Title: US-10-615-659-1_COPY_369_1274
Perfect score: 906
Sequence: 1 gagacatgcacgtcagc.....tcgagctcctgtgagggcag 906

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_hcc.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_ges1.*

9: gb_ges2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	906	100.0	1259	9 AY415398	Homo sapi
2	863.4	95.3	1257	9 AY415399	Pan trogl
3	768	84.8	770	7 CR745100	CR745100
4	608.8	67.2	1265	9 AY415400	Mus muscu
5	608.8	67.2	2720	3 AK029745	Mus muscu
6	608.8	67.2	2969	3 AK030151	Mus muscu
7	398.8	44.0	4184	3 BX648175	Homo sapi
8	378.4	41.8	1023	4 BM808637	AGENCOURT
9	377	41.6	2520	4 AK080321	Mus muscu
10	370	40.8	775	5 BU937372	AGENCOURT
11	360	39.7	1022	4 BM808516	AGENCOURT
12	320.6	35.4	693	4 BM717061	UI-E-EJO-
13	316	34.9	861	7 CK653583	AGENCOURT
14	305.8	33.8	896	6 CD515431	AGENCOURT
15	294.6	32.5	777	7 CK597795	AGENCOURT
16	285.2	31.5	762	6 CD627148	56047432J
17	284.2	31.4	503	7 CK604006	AGENCOURT
18	257.8	28.5	583	7 CN349423	170004243
19	252.8	27.9	514	7 CR746623	CR746623
20	245.4	27.1	320	1 AA144756	mr73D04.r
21	243.4	26.9	680	6 CB854808	UI-CF-EC1
22	222.4	24.5	429	7 W31827	zb94f04.r1
23	219.2	24.2	880	5 BX448465	BX448465
24	213.2	23.5	616	7 CN349424	170006000

25	205	22.6	453	7 W31397	W31397 zb91a01.r1
26	199	22.0	590	1 AL045121	AL045121 DKFZp434M
27	195.6	21.6	484	2 BF830679	BF830679 RC3-HT023
28	192.2	21.2	414	7 CN096245	CN096245 EC2CAA11A
29	191.4	21.1	875	5 BQ718874	BQ718874 AGENCOURT
30	190.8	21.1	3328	3 BC028169	Homo sapi
31	190.4	21.0	651	7 CF999072	CF999072 AGENCOURT
32	186.2	20.6	504	5 BQ327883	MR4-RT004
33	185.6	20.5	965	5 BX431541	BX431541 BX431541
34	185.4	20.5	580	5 BP305902	BP305902 BP305902
35	179.8	19.8	724	4 BG916644	602814292
36	176.6	19.5	498	1 AA551103	nk75c11.8
37	173.6	19.2	488	9 CE760083	tigr-g88-
38	168.8	18.6	608	7 CN349425	CN349425 170005999
39	167.6	18.5	812	6 CD648578	CD648578 AUP_103_C
40	158.8	17.5	630	2 BE865183	BE865183 UI-M-BH1-
41	156	17.2	850	5 BX849970	BX849970 BX849970
42	154	17.0	199	1 AA061487	ml32f01.r
43	151	16.7	507	2 BE837573	RC2-FN009
44	149.4	16.5	468	2 BE827753	RC2-ET001
45	146.6	16.2	977	5 BX431542	BX431542 BX431542

ALIGNMENTS

RESULT 1
LOCUS AY415398 1259 bp DNA linear GSS 12-DEC-2003
DEFINITION Homo sapiens HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION AY415398
VERSION AY415398.1 GI:39771357
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source Location/Qualifiers
1..1259
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1259
/locus_tag="HCM5544"

Query Match 100.0%; Score 906; DB 9; Length 1259;
Best Local Similarity 100.0%; Pred. No. 7e-208;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAGGACATGCACACGTGCGTGGAGGAGCTCTACTGAGCGGAGTGGAGGAC 60
|||||
Db 234 GAGGACATGCACACGTGCGTGGAGGAGCTCTACTGAGCGGAGTGGAGGAC 293

QY	61	CTGACCCAGCAGTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	120
Db	294	CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	353
QY	121	TACTTTTCGAGTGCAGAGCTCTGCTGAATAGAAATACAGTCTGTGTGAACCTTCAGACGGAC	180
Db	354	TACTTTTCGAGTGCAGAGCTCTGCTGAATAGAAATACAGTCTGTGTGAACCTTCAGACGGAC	413
QY	181	ATTGACGGCTCCGGAAACATCTGGATTATTAAGCCGCGCCGCAAGTCCCGGGGCCAGAC	240
Db	414	ATTGACGGCTCCGGAAACATCTGGATTATTAAGCCGCGCCGCAAGTCCCGGGGCCAGAC	473
QY	241	ATAGTGTGCATGACCGCTGTGGAGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	300
Db	474	ATAGTGTGCATGACCGCTGTGGAGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	533
QY	301	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGCGCTGTCTCATCTGTGAC	360
Db	534	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGCGCTGTCTCATCTGTGAC	593
QY	361	ACCAAGTTCGACATCAGACAGTGGTTCTCGTCAAGCTGGAAACCCCTGCACATCTGG	420
Db	594	ACCAAGTTCGACATCAGACAGTGGTTCTCGTCAAGCTGGAAACCCCTGCACATCTGG	480
QY	421	TTCTACAAGGAGTGTACTTGGGTTCTCAACTCAGCGCTTCTCCTGGACAAGCTGGAC	480
Db	654	TTCTACAAGGAGTGTACTTGGGTTCTCAACTCAGCGCTTCTCCTGGACAAGCTGGAC	713
QY	481	AGCGCCATCAGCTGTGCAACACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	540
Db	714	AGCGCCATCAGCTGTGCAACACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	773
QY	541	AGCCCCCTCTGCTCCCGCACACAACATGTGGACACAGCAGGTTCCAGAGTACCTGCGAG	600
Db	774	AGCCCCCTCTGCTCCCGCACACAACATGTGGACACAGCAGGTTCCAGAGTACCTGCGAG	833
QY	601	CGCAGGGCGCTGGCGCGCTGTGGGCGAGCGTCACTACCCCGTCCATGAAGAGGCCATC	660
Db	834	CGCAGGGCGCTGGCGCGCTGTGGGCGAGCGTCACTACCCCGTCCATGAAGAGGCCATC	893
QY	661	GCCACGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGTTC	720
Db	894	GCCACGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGTTC	953
QY	721	TAGGGGCTGACTTCGTCTTGGAGGGACTTCAGGCCCTGCTGATCGAGATCAATTCC	780
Db	954	TAGGGGCTGACTTCGTCTTGGAGGGACTTCAGGCCCTGCTGATCGAGATCAATTCC	1013
QY	781	AGCCCCCACCATGACCCCGTCCAGCCGGTCCAGGCCAGCTGTGTGCACAGGTGCGAGGAG	840
Db	1014	AGCCCCCACCATGACCCCGTCCAGCCGGTCCAGGCCAGCTGTGTGCACAGGTGCGAGGAG	1073
QY	841	GACACCATCAAGTGGCCGTGGACCCGAGTGTGACATCGGCNACTTCGAGCTCTGTGG	900
Db	1074	GACACCATCAAGTGGCCGTGGACCCGAGTGTGACATCGGCNACTTCGAGCTCTGTGG	1133
QY	901	AGGCAG 906	
Db	1134	AGGCAG 1139	
RESULT 2			
LOCUS	AY415399	1257 bp DNA linear	GSS 12-DEC-2003
DEFINITION	Pan troglodytes HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.		
ACCESSION	AY415399		
VERSION	AY415399.1	GI:39771358	
KEYWORDS	GSS.		
SOURCE	Pan troglodytes (chimpanzee)		
ORGANISM	Pan troglodytes		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.		
AUTHORS	1 (bases 1 to 1257) Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.		
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios		
JOURNAL	Science 302 (5652), 1960-1963 (2003)		
PUBMED	14671302		
REFERENCE	2 (bases 1 to 1257) Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA		
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.		
FEATURES	Location/Qualifiers		
Source	1..1257 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1..>1257 /locus_tag="HCM5544"		
gene			
ORIGIN			
Query Match	95.3%;	Score 863.4;	DB 9; Length 1257;
Best Local Similarity	95.9%;	Pred. No. 1.3e-197;	
Matches	867;	Conservative 0;	Mismatches 37; Indels 0; Gaps 0;
QY	1	GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGTGGAGGAC	60
Db	234	GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGTGGAGGAC	293
QY	61	CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	120
Db	294	CTGACCCAGCAGTACTCTCCCTCGTTTCANNNNATGCTTTTCATCTCCAATTCAAGAAAT	353
QY	121	TACTTTTCGACATGCCAGGCTCTGTGAAATAGAAATCACTGTGTGAAACCTTCAGACGGAC	180
Db	354	TACTTTTCGANNNNCCAGGCTCTGTGAAANNNTTCAANNNTNNNNCTNAGACGGAC	413
QY	181	ATTGACGGCTCCGGAAACATCTGGATTATAAGCCGCGCCGCAAGTCCCGGGGCCAGAC	240
Db	414	ATTGACGGGNNCCGAAACATCTGGATTATAAGCCGCGCCGCAAGTCCCGGGGCCAGAC	473
QY	241	ATAGTGTGCATGACCGCTGTGGAGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	300
Db	474	ATAGTGTGCATGACCGCTGTGGAGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	533
QY	301	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGCGCTGTCTCATCTGTGAC	360
Db	534	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGCGCTGTCTCATCTGTGAC	593
QY	361	ACCAAGTTCGACATCAGACAGTGGTTCTCGTCAAGCTGGAAACCCCTGCACATCTGG	420
Db	594	ACCAAGTTCGACATCAGACAGTGGTTCTCGTCAAGCTGGAAACCCCTGCACATCTGG	653
QY	421	TTCTACAAGGAGTGTACTTGGGTTCTCAACTCAGCGCTTCTCCTGGACAAGCTGGAC	480
Db	654	TTCTACAAGGAGTGTACTTGGGTTCTCAACTCAGCGCTTCTCCTGGACAAGCTGGAC	713
QY	481	AGCGCCATCAGCTGTGCAACACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	540
Db	714	AGCGCCATCAGCTGTGCAACACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	773
QY	541	AGCCCCCTCTGCTCCCGCACACAACATGTGGACACAGCAGGTTCCAGAGTACCTGCGAG	600
Db	774	AGCCCCCTCTGCTCCCGCACACAACATGTGGACACAGCAGGTTCCAGAGTACCTGCGAG	833

Qy	601	GCCAGGCGGTGGCGCGTGTGGGGCAGCGTCACTACCGTGCCATGAAGAAGGCATC	660
Db	834		893
Qy	661	GCCACGCCCATGAAGTGGCCACGACCA CGTGGAGCCTCGCAAGAACAGCTTTGAGCTC	720
Db	894		953
Qy	721	TACGGGCTGACTTCGTCCTTGGGAGGAGACTTCAGGCGCTGGCTGATCGAGATCAATTCC	780
Db	954		1013
Qy	781	AGCCCCACCATGCAACC GGTCCA CCGCCCGTCA CCGCCCGTAGCTGTGTGCA CAGGTGCAGGAG	840
Db	1014	AGCCCCACCATGCAACC GGTCCA CCGCCCGTCA CCGCCCGTAGCTGTGTGCA CAGGTGCAGGAG	1073
Qy	841	GACACCATCAGGTGGCGGTGGACCGCAGCTGTGCATCGGCACTTCGAGCTCTCTGG	900
Db	1074	GACACCATCAGGTGGCGGTGGACCGCAGCTGTGCATCGGCACTTCGAGCTCTCTGG	1133
Qy	901	AGGC 904	
Db	1134	AGGC 1137	

RESULT 3					
CR745100					
LOCUS	CR745100	770 bp	mRNA	linear	EST 30-AUG-2004
DEFINITION	CR745100	Soares_testis_NHT Homo sapiens	cDNA clone	IMAGP97102270 ;	
		IMAGE:1645239	5', mRNA sequence.		

EST.	U012000.1	GI:12000.1
KEYWORDS		
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 770)	
AUTHORS	Eberl, L., Heil, O., Hennig, S., Korn, B., Neubert, P., Partsch, E., Peters, M., Radelof, U. and Schneider, D.	
TITLE	I.M.A.G.E. cDNA Clone Collection	
JOURNAL	Unpublished (2004)	
COMMENT	Contact: Inge Arlart	

Heubnerweg 6, D-14059 Berlin, Germany
Email: www.rzpd.de
RZPD; IMAGP37102270.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Contact: Inge Airlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r. Primer sequence: TTTTACACAGGAACAGCTATGAC.

```

1..770
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1645239"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/notes="vector: p7T73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTTACCAATCTCAAGTGGGCGGCCCAATTTTTTTTTTTT 3' ].
Double-stranded cDNA was ligated to Eco RI adaptors

```

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN	Query Match	84.8%	Score 768;	DB 7;	Length 770;
	Best Local Similarity 99.7%;		Pred. No. 1.4e-174;		
	Matches 768; Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;	
Qy	35	ACCTCACTGAGGCGGAGTGGGAGGACTGACCCAGCAGTACTACTCCCTCGTTTCATGGCG	94		
Db	1	ACCTCACTGAGGCGGAGTGGGAGGACTGACCCAGCAGTACTACTCCCTCGTTTCATGGCG	60		
Qy	95	ATGCTTTTCATCTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTCTGCTGAATAGAA	154		
Db	61	ATGCTTTTCATCTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTCTGCTGAATAGAA	120		
Qy	155	TCACGCTGTGAACCTCTCAGACGGACATTCAGCGGCTCCGGACATCTCGAATATTAAGC	214		
Db	121	TCACGCTGTGAACCTCTCAGACGGACATTCAGCGGCTCCGGACATCTCGAATATTAAGC	180		
Qy	215	CCGCGGCCAAGTCCCGGGGCCGAGACATAGTGTGCATGACCGCTGTGGAGGAGATCCTGG	274		
Db	181	CCGCGGCCAAGTCCCGGGGCCGAGACATAGTGTGCATGACCGCTGTGGAGGAGATCCTGG	240		
Qy	275	AGCTGGCAGCTGCAGACCAACCCCTCTTTTCAGGGACAACAAGTGGGTGGTCCAGAAGTACA	334		
Db	241	AGCTGGCAGCTGCAGACCAACCCCTCTTTTCAGGGACAACAAGTGGGTGGTCCAGAAGTACA	300		
Qy	335	TCGAGACGGCGTGCCTACTGTGACACCAAGTTTCACATCAGACAGTGGTTCCTCGTCA	394		
Db	301	TCGAGACGGCGTGCCTACTGTGACACCAAGTTTCACATCAGACAGTGGTTCCTCGTCA	360		
Qy	395	CGGACTGGAAACCCCTTGACCATCTGTTCTACAAGGAGAGTTACTTTCGCGTTTCTCAACTC	454		
Db	361	CGGACTGGAAACCCCTTGACCATCTGTTCTACAAGGAGAGTTACTTTCGCGTTTCTCAACTC	420		
Qy	455	AGCGCTTCTCCTTGGACAAAGCTGGACAGCGCCATCACCCTGTGCACAAACCGCGTCCAGA	514		
Db	421	AGCGCTTCTCCTTGGACAAAGCTGGACAGCGCCATCACCCTGTGCACAAACCGCGTCCAGA	480		
Qy	515	AGTACTGGAAGATGATGTGGGCGCAGCCCTGCTGCCCGCACCAACATGTGACCA	574		
Db	481	AGTACTGGAAGATGATGTGGGCGCAGCCCTGCTGCCCGCACCAACATGTGACCA	540		
Qy	575	GCACCAAGTTTCCAGGAGTACCTGCAGCGCCAGGGCCGTGGCGCGTGTGGGGCAGCGTCA	634		
Db	541	GCACCAAGTTTCCAGGAGTACCTGCAGCGCCAGGGCCGTGGCGCGTGTGGGGCAGCGTCA	600		
Qy	635	TCTACCGTCCATGAAGAAGGCCATCGCCACGCCATGAAGTGGGCCACGAGCACCGTGG	694		
Db	601	TCTACCGTCCATGAAGAAGGCCATCGCCACGCCATGAAGTGGGCCACGAGCACCGTGG	660		
Qy	695	AGCCTCGAAGAACAGCTTTGAGCTCTACGGGGCTGACTTTCGTCCTTGGGAGGACTTCA	754		
Db	661	AGCCTCGAAGAACAGCTTTGAGCTCTACGGGGCTGAGCTTTCGTCCTTGGGAGGACTTCA	720		
Qy	755	GGCCCTGGCTGATCGAGATCAATTCAGGCCCAACCATGACACCGTCCACG	804		
Db	721	NGCCCTGGCTGATCGAGATCAATTCAGGCCCAACCATGACACCGTCCACG	770		

RESULT 4			DNA linear	GSS 12-DEC-2003
AY415400				
LOCUS		1265 bp		
DEFINITION	AY415400 Mus musculus HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.			
ACCESSION	AY415400			
VERSION :	AY415400.1	GI:39771359		
KEYWORDS	GSS.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1265)

REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios

JOURNAL Science 302 (5652), 1960-1963 (2003)

PUBMED 14671302

REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES
Location/Qualifiers
source
1..1265
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1265
/locus_tag="HCM5544"

ORIGIN

Query Match 67.2%; Score 608.8; DB 9; Length 1265;
Best Local Similarity 80.0%; Pred. No. 4e-136;
Matches 734; Conservative 0; Mismatches 172; Indels 12; Gaps 1;

QY 1 GAGGACATCGACACGTCAGACATGCGGTGGAGGACCTCACTGAGCGCAGGTGGAGGAC 60
DB 228 GAAGACATAGACGTATCGGAGCGCAGCAGCGGCTTGTCTGAGGAGGAATGGAATGAC 287

QY 61 CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATCAAGAAAT 120
DB 288 CTGACACACAGTACTATCTACTGTTTCATGGCAATGCTTCCATCACTGATTCGAAAAGT 347

QY 121 TACTTTTCGAGTGCAGAGCTCTGCTGAATAGAAATCAGCTCTGTGAACCTCTGACGGAC 180
DB 348 TACTTTTCGCAATGCCAAGCTCTGCTGAGTAAGATCAGTTCTGTGAACCTCTCAGACAG 407

QY 181 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCGCGCCAAAGTCCCGGGCCGAGAC 240
DB 408 ATCGATGGGATCCGGAACATCTGATCATTAAGCCTGACGCAAGTCCCGGGTCCAGAT 467

QY 241 ATAGTGTGATGACCGGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTT 300
DB 468 ATTGTGTGATGACCGGTGTGGAGAACATCTGTGAGTCTGTGGCCGACAGCCAGACC 527

QY 301 TCCAGGGAACAAGTGGGTGTCAGAGATGATCGAGACGCCGCTGCTCATCTGTGAC 360
DB 528 ACAAGGGAACAACAATGGGTGTCCAGAGATGATCGAGACCCCAATGCTCATCTACGAC 587

QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTGACCATCTCG 420
DB 588 ACCAAGTTTGACATCAGACAGTGGTTCCTCGTCACAGACTGGAATCCCTTAACCATCTGG 647

QY 421 TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGTTCTTCCCTGGACAAGCTGGAC 480
DB 648 TTCTACAAGAAAGCTACCTCCGTTCTCCACACAGGGTTTCTCCCTGGACAAGCTGGAC 707

QY 481 AGGCCCATCCACTGTGCAACAACCGCGTCCAGAGTACTGGAAGATGATGTGGGCCCGC 540
DB 708 AGGCCCATCCACTGTGTAACTCCATCCAGAGGCGCTCAAAAATGATTAAGAGCGC 767

QY 541 AGCCCCCTGCTCCCGCACACACATGTGGACCGACGAGGTTCCAGGAGTACCTGCGAC 600
DB 768 AGTCCCGCTGCTACTTGTCTATAACATGTGGACGAGCACTCGTTTCCAGGAGTACCTGCGAC 827

QY 601 CCCCAGGGCGGTGGCGCCGCTGTGGGGCAGCGTCATCTACCCGTCCATGAAGAAGCCATC 660
DB 828 AAGAGGGGCGGAGGAGAACGTTGGGGTAGCATCATCTACCCGTTATGAAAAGAGCTGTC 887

QY 661 GCCCAGCCCATGAAGTGGCCCGCAGGACCACTGGAGCCCTGCGAAGACAGCTTTGAGCTC 720
DB 888 ACCAATGCCATGAGGTAGCCCGAGGACCACTAGAACCCGTAAGAACAGCTTCGAGCTC 947

QY 721 TAGCGGGCTGACTTCGTCTCTGGAGGAGACTTCAGGCCCTGGCTGATCGAGATCAATTCC 780
DB 948 TAGGAGCGCACTTCTCTCTGGGGCGAGACTTCAGGCCATGGCTTATCGAGATCACTCC 1007

QY 781 AGCCCCCACCATCACCCGTCACGGCGGTACGGCCAGCTGTGTGCA CAGGTGCAAGGAG 840
DB 1008 AGCCCCCACCATCACCCCTCCACTCCCGTCACAGCTCAGCTCTGTGCCCGAGTGCAGGAG 1067

QY 841 GACACCATCAAGGTGGCGG-----TGGACCGCAGCTGTGACATCGGCAACTTC 888
DB 1068 GACACCATCAAGGTGGTGTGTGATCGCAAACTGGACCGAAACTGTGACATCGGCAACTTT 1127

QY 889 GAGCTCCTGTGGAGGCAG 906
DB 1128 GAGCTTCTGTGGCGGAG 1145

RESULT 5
AK029745 2720 bp mRNA linear HTC 03-APR-2004

LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930524K07 product:hypothetical Tubulin-tyrosine ligase containing protein, full insert sequence.

DEFINITION

ACCESSION AK029745 GI:26325681

VERSION AK029745.1

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636

REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED 11042159

REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED 11076861

REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE
5 The FANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

6 (bases 1 to 2720)
Nature 420, 563-573 (2002)
Fukuda, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Adachi, J., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashida, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kage, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, H., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Segabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp).
URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,
Fax: 81-45-503-9216

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>.

Location/Qualifiers

1. 2720

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM:DB:4930524K07"

/db_xref="taxon:10090"

/clone="4930524K07"

/sex="male"

/tissue_type="testis"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

934..2490

/note="unnamed protein product; hypothetical

Tubulin-tyrosine ligase containing protein

(InterPro|IPR004344, evidence: InterPro)

putative"

/codon_start=1

/protein_id="BAC26595.1"

/db_xref="GI:26325682"

/translation="MLNHYGKTSFTKIGLCNMRSLPWVQANPTFPFCVGLCT

ESKQFLDPRFTVAASILKVLVHONYCSKVKGSKKEAKNSDPSPKDPEPDL

KLPSLQVDVDTACKVQAYLQLEHEDIDVSENSTEALSEBENLDLTOQYLLVGN

ASLTSKSYFAQCQALLSKSSNPQTEIDGIRNIWIKPAKSRGDIIVCMRDVENI

LSLVADSSQTKNKVQYKIIETPMIYDTRQWFLVTDWNPDLTIWFKESYLK

FSTQFSLDLSDAIHLCCNNSIQRLKNDKERSPLLPCHNWTSTRFQYLQKRGKG

TWGSIIYPSMKRAVTVNMRVQDHVEARKNSFELYGADFLIGDRDFKPLWLEINSPTM

HPSPPTVAQLCAVOEDTIKVVDRLDRNCIDIGNPELLWQPAVELPPFNGSDLCVE

GISVKKAKQMPPIASVGLSESLDAPPKVRSRALMETVIRPRTTVRDQWRREAK

GFRVISKRR"

ORIGIN

Query Match 67.2%; Score 608.8; DB 3; Length 2720;
Best Local Similarity 80.0%; Pred No. 4.5e-136;
Matches 734; Conservative 0; Mismatches 172; Indels 12; Gaps 1;

1 GAGGACATCGACAGCTCAGCAGATGCGCGTGAGGACCTCACTGAGGCGGAGTGGAGGAC 60

|||||

1318 GAAGACATAGACGATCGGAGGCCAGCCGAGGCGCTTGCTGAGGAGGAATGGAATGAC 1377

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Qy 61 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAGAAAT 120

|||||

Db 1378 CTGACACAGTACTACTCTACTGGTTCATGGCAATGCTTCCCATCTGATTCGAAAGT 1437

|||||

Qy 121 TACTTTTCGAGTCGCCAGGCTCTGCTGAATAGAAATCACTGCTGTGAACCTTCAGACGAC 180

|||||

Db 1438 TACTTTGCGCAATGCGCAAGCTCTGCTGAGTAAGATCAGTTCTGTGAACCTTCAGACGAC 1497

|||||

Qy 181 ATTGACGGGCTTCGGAAACATCTGGATTATTAAGCCCGCGGCAAGTCCCGGGCGGAGAC 240

|||||

Db 1498 ATCGATGGGATCGGAAACATCTGGATCATTAAGCTTCAGCAAGTCCCGGGCTCGAGAT 1557

|||||

Qy 241 ATAGTGTGATGACCGCTGTGGAGGAGATCCTGAGCTGCGAGCTGCGAGACCACTCTT 300

|||||

Db 1558 ATTGTGTATGACCGCTGTGGAGAACATCTGAGTCTGGTGGCCGAGACGACGAC 1617

|||||

Qy 301 TCCAGGACCAACAGTGGTGGTCCAGAACTACATCGAGACGCGCTGCTCATCTGTGAC 360

|||||

Db 1618 ACAAGGACCAACAAATGGTGGTCCAGAACTACATCGAGACGCGCTGCTCATCTGTGAC 1677

|||||

Qy 361 ACCAAGTTGACATCAGACAGAGTGGTCTCGTCAAGCTGGAACCCCTGACCATCTGG 420

|||||

Db 1678 ACCAAGTTTGCATCAGACAGAGTGGTCTCGTCAAGCTGGAATCCCTTAACCATCTGG 1737

|||||

Qy 421 TTCTACAAGGAGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGAC 480

|||||

Db 1738 TTCTACAAGGAGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGAC 1797

|||||

Qy 481 AGCGCCATCCACCTGTGTCACAAACCGCGTCCAGAACTACCTGGAAGATGATGTGGCGGC 540

|||||

Db 1798 AGCGCCATCCACCTGTGTCACAAACCGCGTCCAGAACTACCTGGAAGATGATGTGGCGGC 1857

|||||

Qy 541 AGCCCCCTGTGTCGCGGCAACAACATGTGACACGACACAGGTTCAGGAGTACCTGAC 600

|||||

Db 1858 AGTCCGCTGTCTGCTTGTCTAATCATGTGACACGACCTGTTTCCAGGAGTACCTGAC 1917

|||||

Qy 601 CGCCAGGCGCTGCGCGGTGTGGCGAGCTCATCTACCGCTCCATGGAAGAGGCCATC 660

|||||

Db 1918 AAGAGGCGCGAGGAGGAGGAGCTGGGGTAGCATCATCTACCGCTCATGGAAGAGGCTGTC 1977

|||||

Qy 661 GCCCAGCCATGAAGTGGCGCCAGGACCACTGGAGCTCGCAAGAACAGCTTTGAGCTC 720

|||||

Db 1978 ACCAATGCCATGAGGTAGCCAGGACCACTAGAGCCCGTGAAGACAGCTTCGAGCTC 2037

|||||

Qy 721 TAGCGGCTGACTTTCGCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTC 780

|||||

Db 2038 TAGCGAGCCGACTTCATCTCTGGGCGAGACTTCAAGCCATGGCTTATCGAGATCAATTC 2097

|||||

Qy 781 AGCCCCCACCATGACCCGCTCCAGCGGTACCGCCGAGCTGTGTGACAGGTGCGAGG 840

|||||

Db 2098 AGCCCCCACCATGACCCCTCCACTCCCGTCCAGCTCAGCTCTGTGCGCCAGGTGCGAGG 2157

|||||

Qy 841 GACACCATCAAGTGGCGCG-----TGACCGCAGCTGTGACATCGGCAATTC 888

|||||

Db 2158 GACACCATCAAGTGGTGGTGGATCGCAAACTGACCGAACTGTGACATCGGCAATTC 2217

|||||

Qy 889 GAGCTCTGTGGAGGACG 906

|||||

Db 2218 GAGCTCTGTGGCGGACG 2235

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

RESULT 6

AK030151

LOCUS

DEFINITION

AK030151

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK030151

GI:26326134

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AK030151

Mus musculus adult male testis cDNA, RIKEN linear

library, clone:4933401B17 product: hypothetical Tubulin-tyrosine

ligase containing protein, full insert sequence.

AK030151

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 2969)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
Location/Qualifiers
1. .2969
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"

REFERENCE
source

/db_xref="FANTOM DB:4933401B17"
/db_xref="taxon:10090"
/clone="4933401B17"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
490. .2835

CDS

note="unnamed protein product; hypothetical
Tubulin-tyrosine ligase containing protein
(InterPro: IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26811.1"
/db_xref="GI:26326135"
/translation="MENGERRKSLSTLSDGDKHKKLKGIPQDLSSSPKLDRYKIA
RQLTAKAIKERKIFSIYGHYPVIRLTKRGKVVKKFPPKALQNLGSSDKSAETKE
NOEIALERFDDLDHDMVRLVKNKEIPLYLWTKRDVVDVHSLTCDQMLNHYGKTASFTT
KIGLCNMRSLPWYVOAHPNTFFPKCYGLTSEKQEFUDDFRRTVAASILKWWLJHQ
NYCSVKGSKKEEANDSPSPKDPNPDLKPSLGQVVDTKVCOQALSKISVNPQT
DIDVSEASTALSEBENWDLTQYVLLVHGNASITDSKSYFAQCQALLSKISVNPQT
EIDGRNTWIIKPAASKRGDIVCDRVENILSLVAADSOTTCKDNKWWQVKYIETPLK
IYDTKEDIQWFLVTDWNPLOTIFVKESVLRSTORFSLDKLSDAHLNNSIQRLLK
NKNRSPFLPCHNMTSTRFQYLRQKRGRTGWSIILPSPMKRAVTNMRVQDHVEA
RNXSPFLGADFLGRDFKPLIEINSSFTMPSTPVAQLCAKQAOEDTIKVVVDRL
DRNCIDIGNFELLWRQPAVELPFGNSDLCEGISVKKAKKQMPPTASVGLSESLIDAP
PKVRSARALMETVIRPRTTVRQDWKBEAKVLSTWSPMVMDAEVGRAKFIYAFEV
NDYQHDVNDKSHKSGYRVQSSKVPGLTISAQHPALFAQTMKPTQWTSPPPTASGNH
RDSPPFCPIVFEELWLHPNSQRRPSSCILQSRAGWIRGIP"

ORIGIN

Query Match 67.2%; Score 608.8; DB 3; Length 2969;
Best Local Similarity 80.0%; Pred. No. 4.6e-136;
Matches 734; Conservative 0; Mismatches 172; Indels 12; Gaps 1;

QY 1 GAGGACATCGACAGTCAGCAGATGCGGTGGAGGACCTCAGTGAGCCGAGTGGGAGGAC 60
DB 1315 GAAGACATAGACGTATCGGAGGCCAGCACCGAGGCGCTTGCTGAGGAGGAATGGAATGAC 1374
QY 61 CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAGAAAT 120
DB 1375 CTGACACAGCAGTACTATCTACTGTTTCATGGCAATGCTTCCATCACTGATTCGAAAAGT 1434
QY 121 TACTTTTCGAGTGGCAGGCTCTGCTGAATAGAATCACGTTGTGAACCTCAGACGGAC 180
DB 1435 TACTTTGCGCAATGCCAAGCTCTGCTGAGTAAGATCAGTTCTGTGAACCTCAGACAGAG 1494
QY 181 ATTGACGGGCTCGGACACATCTGGATTATAAGCCCGGCGCAAGTCCCGGGCCGAGAC 240
DB 1495 ATCGATGGGATCCGAACATCTGGATCATATAAGCCCTGCGAGCAAGTCCCGGGGTCGAGAT 1554
QY 241 ATAGTGTGCATGCAGCCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT 300
DB 1555 ATTGTTGTATGGACCGTGTGGAGAACATCTGTAGTCTGTGGCCGCGACAGCCAGACC 1614
QY 301 TCCAGGGACAAACAGTGGGTGTGTCAGAAAGTACATCGAGACCGCGCTGCTCATCTGTGAC 360
DB 1615 ACAAGGACAAACAAATGGGTGTGTCAGAAAGTACATCGAGACCCCAATGCTCATCTACGAC 1674
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCATCTGG 420
DB 1675 ACCAAGTTTGATCAGACAGTGGTTCCTCGTGTGTCAGAGTGGAAATCCCTTAACCATCTGG 1734
QY 421 TTCTCAAGGAGAGTTACTTTGGGTTTCTCAACTCAGCGCTTCTCCTCGACCAAGCTGGAC 480
DB 1735 TTCTACAAAGAAAGTACTCTCGCTTCTCCACACAGCGCTTCTCCTGCAAGCTGGAC 1794
QY 481 AGCGCCATCCACTGTGCAACAAACCGCTGCAGAAAGTACCTGAAAGATGATGTGGGCCCGC 540
DB 1795 AGCGCCATCCACTGTGTAACAACTCCATCCAGAGCGCGCTCAAAAATGATAAAGAGCGC 1854
QY 541 AGCCCCCTGCTGCCCGCACACAAATGTGGACAGCACAGGTCACGAGTACCTCGAC 600

```
Db 1855 AGTCGGCTGTACCTTGTATACATGTGGACAGCACTCGTTTCCAGGAGTACCTGCAG 1914
Qy 601 CGCCAGGGCGGTGGCGCGTGTGGGCGAGCGTCACTACCGCTCCATGAGAAGCGCATC 660
Db 1915 AAGAGGGGCGGAGGAGAACTGGGGTAGCATCATCTACCGCTCTATGAAAAGAGCTGTC 1974
Qy 661 GCCACGCGCATGAAGGTGGCGGAGGACCACTGGAGCTCCCAAGAACAGCTTTGAGCTC 720
Db 1975 ACCAATGCCATGAGGTAGCCAGGACCACTAGAGCCGCTAGAACAGCTTCGAGCTC 2034
Qy 721 TACGGGGCTGACTTTCGCTCTTTGGAGGAGGACTTCAGGCGCTCGCTGATCGAGATCAATTC 780
Db 2035 TACGAGCGGACTTTCATCTCGGGCGGAGACTTCAAGCCATGGCTTATCGAGATCAATCC 2094
Qy 781 AGCCCCACCATGCAACCGGTCAACCGGTCAGCGCCAGCTGTGTGACAGGTCCAGAG 840
Db 2095 AGCCCCACCATGCAACCGGTCACCTCCGCTCAAGCTCAGCTCTGTGCGCCAGGTCCAGAG 2154
Qy 841 GACACCATCAAGGTGGCGG-----TGACCGCGAGCTGTGACATCGGCAACTTC 888
Db 2155 GACACCATCAAGGTGGGTGGTGTGATCGCAACTGGACCGAACTGTGACATCGGCAACTTT 2214
Qy 889 GAGCTCTCTGTGGAGGCGAG 906
Db 2215 GAGCTCTGTGGCGGCGAG 2232

RESULT 7
HSM808322 4184 bp mRNA linear HTC 22-SEP-2004
LOCUS Homo sapiens mRNA; cDNA DKFZp686D076 (from clone DKFZp686D076).
ACCESSION BX648175
VERSION BX648175.1 GI:34367334
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4184)
AUTHORS Ansoorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
CONSTRMT The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686D076) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686D076
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.

FEATURES
source 1..4184
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp686D076Q"
/db_xref="taxon:9606"
/clone="DKFZp686D076"
/tissue_type="cervix"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
DH10B; sites SfiI + SfiIb"
/dev_stage="adult"
/note="hypothetical protein"
1..4184
/gene="DKFZp686D076"
2529..3587
/cds
/gene="DKFZp686D076"
```

```
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH10554.1"
/db_xref="GI:50949619"
/tranlation="MDIDKDLLEAPLYLTPEWSLFLQRYQYVHVEGAEALRHLDTVQVR
CEDILQOLQAVVPQIDWEGDRNTWIKPGAKSRGIMCMHLEMLKLVNGNPNVNMK
DGKVVQKYIERPLLIIFGTFDLRQWPLVTDNPLTVWFYRDSVIRFSTQPSFSLNLD
NSVHLNNSLOKHELENSCHRHPLLPDNNWSSQRFQAHLOEMGAPNASTIIVPMKD
AVTHALQTSQDTVQCRKASFELYGDFVFGEDFQPLWLEINASTMAPSTAVTARLCA
GQADTLRVVVIDRLNRNCDTGAFELIYKQPVTTSPASTPGPSPCLLPYSDTRASSD
DSTASWALRPCRQARP"

ORIGIN
Query Match 44.08; Score 398.8; DB 3; Length 4184;
Best Local Similarity 66.3%; Pred. No. 2.4e-85;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

Qy 3 GGACATCGACACGTCAGCAGATCCGCTGGAGGACCTCACTGAGGCCGAGTGGAGGAGCCT 62
Db 2531 GGACATCGACAGGACTGAGGCGCCCGCTGTACTCACCCCGAGGGCTGCTCCCTTT 2590
Qy 63 GACCCAGCAGTACTCTCCCTCGTTATGCGGATGCTTTTTCATCTCCAAATCAAGAAATTA 122
Db 2591 CCTCCAGCGCTACTTACCAAGTGGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCA 2650
Qy 123 CTTTTGGCAGTGCAGGCTCTGCTGTAATAGATCAGCTCTGTGAACCTCAGACGGACAT 182
Db 2651 GGTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCTGGTACCCAGATAGACAT 2710
Qy 183 TGACGGGCTCCGGAACATCTGGATATTAAAGCCGCGGCAAGTCCCGGGCCGAGACAT 242
Db 2711 GGAAGGGGATCGCAACATCTGGATCTGTAAGCCAGGAGCCAGTCCCGCGGACAGGCGAT 2770
Qy 243 AGTGTGATGAGCCGCTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTTC 302
Db 2771 CATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTAT 2830
Qy 303 CAGGGAACAAAGTGGTGGTTCAGAAAGTATACATCGAGACCGCGCTGCTCATCTGTGACAC 362
Db 2831 GAAGGACGGCAAGTGGTGGTGCAGAAAGTATATTGAGCGGCCCTCTCTCATCTTTTGGCAC 2890
Qy 363 CAAGTTTCAGATCAGACAGTGGTTCCTCGTCAAGCTCGGACTGGAACCCCTGACCATCTGTT 422
Db 2891 CAAGTTTGACTCAGACAGTGGTTCCTGTAAGTCTGTAAGTCCAGTCCAGTCCAGTCCAGT 2950
Qy 423 CTAAAGGAGAGTACTTTCGCTTCAACTCAGCGCTTCTCCCTGGAAGCTGGACAG 482
Db 2951 CTACCGGACAGCTATATCCGCTTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTGGACAA 3010
Qy 483 CGCATCCACTGTGCAACAAACCGCTCCAGAGTACCTGGAAGATGATGTGGGCCCGCAG 542
Db 3011 CTCAGTGCACTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCCATCGGCA 3070
Qy 543 CCCCTGCTGCGCGCACACAACTGTGGACCAAGCTGAGTCCAGGAGTACCTGCGAGCG 602
Db 3071 TCCACTGCTTCCCGCAGACAACTGTTGTAGCAGAGGTTCCAGGCCCCCTTGCAGGAG 3130
Qy 603 CCAGGCGCTGGCGCGCTGTGGGCGCAGCGTCACTACCCCTCCATGAAGAGGCGCATCGC 662
Db 3131 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCGTCCCTGGCATGAAGGATGCTGTGAT 3190
Qy 663 CCACGCGCATGAAGGTGGGCCAGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTA 722
Db 3191 CCACGCACTTCAGACCTCCCGAGGACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTA 3250
Qy 723 CGGGGCTGACTTCTGCTTGGGAGGAGCTTCAGGCGCTGCTGATCCAGATCAATTCAG 782
Db 3251 TGGCGCTGACTTCTGTTTGGGAGGAGACTTCCAGGCCCTGGCTGATTGAGATCAACGCCAG 3310
Qy 783 CCCACCATGCAACCCCGTCCACGCGGTCAGCGGCCAGCTGTGTGCAAGGTCGAGGAGGA 842
Db 3311 CCCACGATGGCACCCCTCCACAGCAGTCACTGCGCGGCTCTGTGCTGGCGTGCAGAGCTGA 3370
```


RESULT 10
BU937372 775 bp mRNA linear EST 18-OCT-2002
LOCUS
DEFINITION AGENCOURT 10519180 NIH_MGC_169 Mus musculus cDNA clone
IMAGE:6705256 5', mRNA sequence.
ACCESSION BU937372 GI:24126191
VERSION
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM3002 row: 9 column: 16
High quality sequence stop: 569.
FEATURES
Location/Qualifiers
1..775
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6705256"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_169"
/note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI
(ggccattggcc); Site 2: SfiI (ggcgctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGACAGTGGCCATTACGCCGG-3' and
5'-ATTCTAGAGCCGAGCGGCCGACATG-dT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
ORIGIN
Query Match 40.8%; Score 370; DB 5; Length 775;
Best Local Similarity 81.2%; Pred. No. 1.7e-78;
Matches 446; Conservative 0; Mismatches 91; Indels 12; Gaps 1;
QY 370 GACATCAGACAGTGGTTCTCTCGTCAGGACTGGAACCCCTGACCATTCTGGTTCTACAAG 429
DB 1 GGCATCAGACAGTGGTTCTCTCGTCAGACTGGAATCCCTACCACTCTGGTTCTACAA 60
QY 430 GAGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATC 489
DB 61 GAAAGCTACTCGCTTCTCCACACAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATC 120
QY 490 CACCTGTGMAACACCGCTGCGAAGTACCTGAAGAATGATGTGGCGCGCAGCCCCCTG 549
DB 121 CACCTGTGTAACAACATCCATCCAGAGCGCCTCAAAAATGATAAAGAGCGCATCCGCTG 180
QY 550 CTGCGCGCACACAACATGTGGACACACACAGGTTCCAGAGTACCTGCGAGCGCCAGGGC 609
DB 181 CTACCTGTGTCATACATGTGGACACAGCATCTGTTTCCAGAGTACCTGCGAAGAGGGC 240
QY 610 CGTGGCGCGCTGTGGGCGAGCGTCACTACCCCGTCCATGAAGAAGCCCATCGCCACGCC 669
DB 241 CGAGGAGGAACGTGGGGTAGCATCTACCCCGTCTATGAAAAGAGCTGTCAACCAATGCC 300
QY 670 ATGAAGGTGCCCGACACACCGTGGAGCCCTCGCAAGACAGCTTTCAGCTCTACGGGCT 729
|||||

DB 301 ATGAGGGTAGCCAGGACCAACGCTAGAACCCCGTAAGAACAGCTTTCAGCTCTTACGGAGCC 360
QY 730 GACTTCGCTCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCGCCACC 789
DB 361 GACTTCATCTGNGGCGAGACTTCAAGCCATGGCTTATCGAGATCACTCCAGCCGCCACC 420
QY 790 ATGCACCCCTCCACGCCGCTCAGCGCCAGCTGTGTGCAAGGTGCGAGGAGCACCATC 849
DB 421 ATGCACCCCTCCACTCCGCTCAGCTCTGTGCCAGGTGCGAGGAGCACCATC 480
QY 850 AAGTGGCCG-----TGGACCCAGCTGTGACATCGGCAACTTCGAGCTTCCTG 897
DB 481 AAGGTGGTGGTGGATCGCAAACTGGACCGAAATGTGACATCGGCAACTTTGAGCTTCTG 540
QY 898 TGGAGGCGAG 906
DB 541 TGGCGGCGAG 549
BM808516 1022 bp mRNA linear EST 05-MAR-2002
AGENCOURT_6617783 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734160
5', mRNA sequence.
ACCESSION BM808516
VERSION
KEYWORDS EST.
SOURCE BM808516.1 GI:19125339
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12739 row: a column: 09
High quality sequence start: 78
High quality sequence stop: 658.
FEATURES
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734160"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."
ORIGIN
Query Match 39.7%; Score 360; DB 4; Length 1022;
Best Local Similarity 68.8%; Pred. No. 4.6e-76;
Matches 495; Conservative 0; Mismatches 225; Indels 0; Gaps 0;
QY 129 GCAGTGCAGGCTCTCTGAATAGATCAGCTCTGTGAACCTTCAGACGACATTACCG 188
DB 40 GCGCTGTGAGGACATCTCTGCAGAGCTGCAGGCCGGTATCCCATAGATAGATGAAGG 99
|||||

```
QY 189 GCTCCGGAACATCTGGATTAAAGCCGCGCCCAAGTCCCGGGCGGAGACATAGTGTG 248
Db 100 GGATCGGAACATCTGGATCGTGAAGCAGGAGCCCAAGTCCCGGGCGGAGACATAGTGTG 159
QY 249 CATGGACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCACTCTTTCACAGGA 308
Db 160 CATGGACCACTCGAGGAGATCTGTAAGCTGGTGAACGGCAACCCGCTGGTGAAGGA 219
QY 309 CAAACAGTGGTGTCTCAGAAATACATCAGACGCCGCTGTCTCATCTGTGACACCAAGTT 368
Db 220 CGGCAAGTGGTGTCTCAGAAATATATTGAGCGGCCCTCTCTCATCTTTGGCACCAGTT 279
QY 369 CGACATCAGACAGTGTCTCTCGTCAGGACTGGAACCCCTGACCATCTGTTCTACAA 428
Db 280 TGACCTCTTACAGTGGTCTCTGTAACCTGACTGGAACCCCTTACCCTGGTGTCTACCG 339
QY 429 GGAGAGTTACTTGGGGTCTTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCAT 488
Db 340 CGACAGCTATATCCGCTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTGGACACTCAGT 399
QY 489 CCACCTGTGCAACAAACCGGTCAGAAATACCTGGAAGATGATGTGGCGCGACGCCCT 548
Db 400 GCACCTGTGCAACAACTCCATCCAGAGCACCTGGAGAACCTCATGCCATCGGCATCCACT 459
QY 549 GCTGCCCGCACACACATGTGGACAGCACCAAGCTTCCAGAGTACCTGCAGCGCCAGGG 608
Db 460 GCTTCCCGCACACACATGTGGTGTAGCCAGAGTTCAGGCCACCTGCAGGAGATGGG 519
QY 609 CCGTGGCGCGTGTGGGCGAGCGTCATCTACCCCTCATGAAGAGCCATGCGCCACGC 668
Db 520 TGCCCCAAATGCTTGGTCCACCATCATCTGCTGGCATGAGATGCTGTGATCCAGC 579
QY 669 CATGAAGTGGCCAGACACCACTGGAGCTTCGCAAGACAGCTTTCAGCTTACGGGG 728
Db 580 ACTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGCCAGCTTTGAGCTCTATGGCG 639
QY 729 TGACTTCGCTTGGGAGGACTTTCAGGCCCTGCTGATCGAGATCAATTCAGCCCCAC 788
Db 640 TGACTTCGCTTGGGAGGACTTTCAGGCCCTGCTGATTCAGATCAACCGCCGCCAC 699
QY 789 CATGACCCCGTCCACCGCGGTTCAGGCCAGCTGTGTGCACAGTGCAGGAGGACACCAT 848
Db 700 GATGGCACCCCTCACAGAGTCACTGGCGCGCTCTGTGCTGGCTGCAAGCTGACACCT 759
```

```
RESULT 12
BM717061
LOCUS BM717061 693 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-EJ0-ahk-e-05-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone
BM717061
VERSION UI-E-EJ0-ahk-e-05-0-UI 5', mRNA sequence.
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
```

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

source

1..693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-ahk-e-05-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJ0"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes,
AGATACAGA; lens, CGATTAGCGA; eye anterior segment,
AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG;
Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
library was created for the program, Gene Discovery in the
Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 35.4%; Score 320.6; DB 4; Length 693;
Best Local Similarity 69.1%; Pred. No. 1.4e-66;
Matches 472; Conservative 0; Mismatches 199; Indels 12; Gaps 2;
QY 235 CGAGACATAGTGTGACCGGTGTGGAGAGATCTCTGGAGCTGGCAGCTGCAGACCC 294
Db 1 CGAGGCATCATGTGTCATGACACCTGGAGAGATGCTGAAGCTGGT-GAACGCGACCCC 59
QY 295 CCTCTTTCAGGACAAACAGTGGTGGTCCAGAAATACATCGAGACCGCGCTGCTCATC 354
Db 60 GTGGTGTATGAAGACGCGCAAGTGGTGGTGCAGAAATATATTGAGCGGCCCTCTCATC 119
QY 355 TGTGACACCAAGTTCGACATCAGACAGTGGTCTCTGTACGGAGCTGGAAACCCCTGACC 414
Db 120 TTTGGCACCAGTTTGACCTCAGACAGTGGTCTCTGTAACTGACTGGAAACCCACTTACC 179
QY 415 ATCTGGTTTCAAGGAGAGTATTCTTGGGTCTCAACTCAGCGCTTCTCCCTGGACAAG 474
Db 180 GTGTGGTTTACCGGACAGCTATATCCGCTTTTCCAGCAGGCCCTTCTCCCTGGAAGAC 239
QY 475 CTGGACAGCGCCATCCACCTGTGCAACACCGCGTCCAGAAATACCTGGAAGATGATGTG 534
Db 240 CTGGACAACCTCAGTGCACCTGTGCAACACTCCATCCAGAAGCACCTGGAGAATCATGC 299
QY 535 GGGCGGAGCCCCCTGTGCCCCGACACAAATGTGGACACGACCGAGTTCCAGGAGTAC 594
Db 300 CATCGGCATCCACTGCTTCCGCCAGACAAATGTGGTCTAGCCAGAGGTTCCAGGCCAC 359
QY 595 CTCGAGCGCCAGGCGCGTGGCGCGTGTGGGAGGCTCATCTACCGCTCCATCGAAGAAG 654
Db 360 CTCGAGAGATGGTGGCCCCAAATGCTTGTTCACCAATCATCTGCTGCTGGCATGAGAT 419
QY 655 GCCATCGCCACCGCCATGAGAGTGGCGCCAGGACCACTGGAGGCTTCGCAAGAAACAGCTTT 714
Db 420 GCTGTGATCCACGCACTTCAGACCTCCAGGACACCGGTGCGAGTGTGCGAAGGCCAGCTTT 479

QY 715 GAGCTCTACGGGCTGACTTCCTTGGTGGAGGGAGACTTCAGGCCCTGGCTGATCGAGATC 774
Db 480 GAGCTCTATGGGCTGACTTCCTTGGTTCGGGGAGGAGACTTCAGGCCCTGGCTGATCGAGATC 539
QY 775 AATTCAGCCCCCAGCATGACCCGTCACCCCGGTACCGGCCAGCTGTGTGCACAGGTG 834
Db 540 AACGCCAGCCCCAGCATGGGACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGTGGCGTG 599
QY 835 CAGGAGGACAC-----CATCAAGGTGCCGTGGAGCCGAGCTGTGACATCGGCA 883
Db 600 CAAGCTGAACCTGCGCGGTGTCAATGACCGGATGCTGGACCCGAACTGTGACACAGGAG 659
QY 884 ACTTCGAGCTCTGTGGAGCAG 906
Db 660 CCTTTGAGCTCATCTATAAGCAG 682

RESULT 13
CK653583 861 bp mRNA linear EST 29-JAN-2004
AGENCOURT 17673480 NIH MGC 237 Rattus norvegicus cDNA clone
IMAGE:7113394 5', mRNA sequence.

CK653583
CK653583 GI:41389106
EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 861)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLNL4985 row: e column: 08

High quality sequence stop: 304.

Location/Qualifiers

FEATURES

source

1..861

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="IMAGE:7113394"

/issue_type="testis, pooled"

/lab_host="DH10B Tona"

/clone_lib="NIH MGC 237"

/note="Organ: testis; Vector: pExpress-1; Site 1: EcorV;

Site 2: NotI; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer:

5'-pGACTAGTTCTAGATCGAGCGGCCGCC(NT)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection 1.4kb resulted in an average insert size of 2.4 kb. This primary library is not normalized (normalized primary library is NIH MGC 238) and was constructed by Express Genomics (Frederick, MD)"

ORIGIN

Query Match 34.9%; Score 316; DB 7; Length 861;

Best Local Similarity 74.2%; Pred. No. 1.9e-65;

Matches 439; Conservative 0; Mismatches 150; Indels 3; Gaps 3;

QY 227 CCGGGGCCGAGACATAGTGTGCATGGACCGTGTGGAGAGATCCTCGAGCTGGCAGCTG 286
Db 12 CCGGGATTGGAGATATTGTGTGCATGGACCGTGTGGAGAACATCCTGGATCTGTGTGGCC- 70
QY 287 CAGACCAACCTCTTTTCCAGGGACAACAAGTGGGTGTGCAGAAAGTACATCGAGACGCCGC 346
Db 71 CAGACAGGAGAGACCAAGGAGGACAACAATGGGTGTGCATAGTACATCGAGACCCCGA 130
QY 347 TGTCTATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTTCCTCGTCAGGAGCTGGAAC 406
Db 131 TGTCTATCTATGACCAAGTTTTCGACATCAGACAGTGGTTTCTTGTACAGACTGGAATC 190
QY 407 CCTGACCATCTGGTTTCTACAAGGAGAGTTTACTTTGGCGTTTCTCAACTCAGCGCTTCTCCC 466
Db 191 CCTTAACCATCTGGTTTCTACAAGAGAGACTACCTCCGTTTCTCCAGCGCGCTTCTCCC 250
QY 467 TGGACAAGCTGGACAGCGGCATCCACTGTGTGCACAACCGCGTCCAGAGTACTCTGAAGA 526
Db 251 TGGACAAACTGGACAGCGCTTCCACCTGTGTAACTCCGTCCAGAGCGTGGCATAA 310
QY 527 ATGATGTGGGCGCAGCCCTCGCTGCCCGCACACAACATGTGGACCCAGCAGCAGGTTC 586
Db 311 ATGACAAAAACGAAGTCCACTGTGACATTAACATGTGGACCATCACCCTGTTCC 370
QY 587 AGGAGTACTGACAGCCAGGGCCGTGGCGCGCTGTGGGGCAGCGTCATCTACCCGTCCA 646
Db 371 ATGAGTACTGTGANTATAGGGCCGAGGATGACATGGGGTAGTATCACTACCCGTCTA 430
QY 647 TGAAGAAGCCCATCGC-CCACGCCATGAAGGTGGCCAGGACCAACATGTGAGCGCTCGCAAG 705
Db 431 TGAAGAAGAGCTGTCACTCAACGTGATGTGGTGGCCCAAGATCTTGTGAAGCCCGTAAG 490
QY 706 AACAGCTTTGAGCTCTACGGGCTGACTTCCTCTGGGAGGAGCTT-CAGGCGCTGGCT 764
Db 491 ATTATCTTCGATCTCTATGGAGCTGACTTTAACTCTGGGCGTGAATATAGTCTCTGTGT 550
QY 765 GATCGAGATCAATTCCAGGCCCAACATGACCCCGTCCAGCCCGGTCCAGCGCC 816
Db 551 CTTCAAGATCAACTTCTACTCTGCTTGCACCTCTACTTTGTAAAMACC 602

CD515431 896 bp mRNA linear EST 06-JUN-2003
AGENCOURT 14364765 NIH MGC 181 Homo sapiens cDNA clone
IMAGE:30407505 5', mRNA sequence.

CD515431
CD515431 GI:31447149

EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 896)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Michael Brownstein

cDNA Library Preparation: Invitrogen Corp

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Place: NDM488 row: p column: 10

High quality sequence stop: 678.

Location/Qualifiers

FEATURES

```
source
1. .896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30407505"
/tissue_type="White Matter"
/dev_stage="Unknown"
/lab_host="NIH_MGC_181"
/clone_lib="NIH_MGC_181"
/notes="Vector: pCMV-Sport6.1; Site 1: NotI; Site 2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(Invitrogen). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 33.8%; Score 305.8; DB 6; Length 896;
Best Local Similarity 70.9%; Pred. No. 5.5e-63;
Matches 406; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 304 AGGAGACAAGTGGTGTCCAGAGTACATCAGACGCGGTGCTCATCTGTGACACC 363
DB 25 AAGGACGCGCAAGTGGTGTCCAGAGTATATTGAGCGGCGCCCTCTCATCTTTGGCACC 84
QY 364 AAGTTCCGACATCAGACAGTGGTCTCTCGTCACGACTGGAACCCCTGACCATCTGCTTC 423
DB 85 AAGTTGACCTCAGACAGTGGTCTCTCGTAACTGACTGGAACCCACTTACCGTGTGCTTC 144
QY 424 TACAAGGAGAGTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGC 483
DB 145 TACCGGACAGCTATATCCGCTTTTCCACGAGCCCTTCTCCCTGAAGAACCTGGACAAAC 204
QY 484 GCCATCAGCTGTGCAACAACGCGCTCCAGAGTACCTGGAAGATGATGTGGGCGCGCAGC 543
DB 205 TCAGTGACCTGTGCAACAACCTCCATCCAGAAGCACCCTGGAGAACTCATGCCATCGGCAT 264
QY 544 CCCCTGCTGCCGCGCACACAATGTGACAGCAGCAGCTTCCAGGAGTACCTGACGCGC 603
DB 265 CCATGCTGCTCCGCGCAGACAATGTGTCTAGCCAGAGGTTCAGGCGCCACCTGCGAGG 324
QY 604 CAGGCGCGTGGCGCGGTGTGGGCGAGCTCATCTTACCCGCTCCATGAAGAAGGCCATCGCC 663
DB 325 ATGGGTGCCCAATGCTTGGTCCACCATCATCTGTGCTGGCATGAAGGATGCTGTATC 384
QY 664 CAGCCATGAAGTGGCCAGGACCACTGAGAGCTTCGCAAGAACACTTTGAGCTCTAC 723
DB 385 CAGGCACTTCAGACCTCCAGGACACCGTGAATGTCCGAAGGCCAGCTTTGAGCTCTAT 444
QY 724 GGGGCTGACTTCGTCCTTTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATCCAGC 783
DB 445 GGGGCTGACTTCGTCCTTTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATCCAGC 504
QY 784 CCCACCATGACACCCGCTCCAGCGCGCTCAGCGCCAGCTGTGTGACAGGTGCGAGGAGGAC 843
DB 505 CCCACGATGACACCCCTCCAGCAGTCACTGCGCGGCTCTGTGCTGGCTGCAAGCTGAC 564
QY 844 ACCATCAAGTGGCGCTGGACCGCAGCTGTGAC 876
DB 565 ACCCTGCGCGTGGTCAATTGACCGGAGGCTGGAC 597
```

```
RESULT 15
CK597795
LOCUS 777 bp mRNA linear EST 22-JAN-2004
DEFINITION AGENCOURT 17837841 NIH_MGC_238 Rattus norvegicus cDNA clone
IMAGE:7135519 5', mRNA sequence.
CK597795
VERSION CK597795.1 GI:41110910
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```

```
Rattus.
1 (bases 1 to 777)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: gspbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM15042 row: 0 column: 05
High quality sequence stop: 693.
```

```
FEATURES
Location/Qualifiers
source
1. .777
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7135519"
/tissue_type="testis, pooled"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_238"
/notes="Organ: testis; Vector: pExpress-1; Site 1: EcoRV;
Site 2: NotI; RNA obtained from testis tissue of 8 wk old
animal. Tissues were snap-frozen and kept at -80C before
RNA extraction and purification (Tri-reagent method). cDNA
was primed using oligo-dT primer:
5'-pGACTAGTTAGATCGAGCGCGCCCTTGTCCGAGGAGGAATGGAACGAC
the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
resulted in an average insert size of 1.9 kb. This primary
library is normalized (non-normalized primary library is
NIH_MGC_237) and was constructed by Express Genomics
(Frederick, MD)"
```

ORIGIN

```
Query Match 32.5%; Score 294.6; DB 7; Length 777;
Best Local Similarity 80.0%; Pred. No. 2.7e-60;
Matches 359; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1 GAGGACATCGACAGCTCAGCAGATGCGGTGGAGGACCTCACTGAGGCGGAGTGGAGGAC 60
DB 252 GAGGACATCGACCTGTCCAGACGCGCGGCGCTTGTCCGAGGAGGAATGGAACGAC 311
QY 61 CTGACCCAGAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCAATTTCAAGAAAT 120
DB 312 CTGACACAAACAATACTACTTGTCTGTTTCATGGCAATGCTTCCATCAGGATTCGAGAAT 371
QY 121 TACTTTTCGAGTGCCAGGCTCTGCTGAATAGAAATACAGTCTGTGAACCTCTCAGACGAC 180
DB 372 TACTTTTCGAGTGCCAGGCTCTGCTGAATAGAAATAGCTCTGTGAACCTCTCAGACGAC 431
QY 181 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGGGCGGAGAC 240
DB 432 ATTGACGGGATACGGAACATCTGATCATTAAGCTTCAGCAAGTCCCGGGGCGGAGAT 491
QY 241 ATAGTGTGATGACCGTGTGGAGGAGATCCTGGAGCTGGAGCTGAGACACACCTCTT 300
DB 492 ATTGTGTGATGACCGTGTGGAGAACATCTGGATCTGGTGGCCACAGACGACGAC 551
QY 301 TCCAGGACAAACAAGTGGGTGGTCCAGAGTACATCGAGACGCGCTGCTCATCTGTGAC 360
DB 552 ACCAAGGACAAACAATGGGTGGTCCAGAGTACATCGAGACCCCGATGCTCATCTATGAC 611
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTGACCATCTGG 420
DB 612 ACCAAGTTTGACATCAGACAGTGGTTCCTTGTGTACAGACTGGAAATCCCTT-AACATCTGG 670
```

Qy 421 TTCTACAAGGAGAGTTACTTGCGGTTC 449
|||
Db 671 TTCTACAAGGAGAGTTACTTGCGGTTCAC 699

Search completed: September 25, 2005, 01:56:03
Job time : 3483.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 01:51:04 ; Search time 549.716 Seconds
(without alignments)
9756.464 Million cell updates/sec

Title: US-10-615-659-1_COPY_369_1274
Perfect score: 906
Sequence: 1 gaggacatcgacagtcagc.....tcgagctctctgtggagcag 906

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Geneseqn16Dec04:*
2: Geneseqn1980s:*
3: Geneseqn1990s:*
4: Geneseqn2000s:*
5: Geneseqn2001as:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	906	100.0	1838	ADJ93357	Adj93357 Human BGS
2	906	100.0	3554	ADJ93364	Adj93364 Human tub
3	806.8	89.1	1859	ADJ93362	Adj93362 Human BGS
4	762	84.1	3465	ADJ93363	Adj93363 Human BGS
5	760.4	83.9	1939	ADJ93361	Adj93361 Human BGS
6	726	80.1	726	ADJ93359	Adj93359 Human BGS
7	427	47.1	490	ACH91699	Ach91699 Human gen
8	427	47.1	101270	ADQ17814	Adq17814 human sof
9	400.4	44.2	2881	AAC77214	Aac77214 Human ORF
10	398.8	44.0	2326	AAH16735	Aah16735 Human cDN
11	398.8	44.0	2380	AAS99894	Aas99894 Polynucle
12	398.8	44.0	2538	ADS73168	Ads73168 Human kid
13	398.8	44.0	2553	ACN37881	Acn37881 Tumour-as
14	398.8	44.0	2848	ADM03081	Adm03081 Human cDN
15	398.8	44.0	3001	AAI58606	Aai58606 Human pol
16	398.8	44.0	3001	ADQ98824	Adq98824 DNA encod
17	398.8	44.0	3001	ADB48584	Adb48584 Novel hum
18	398.8	44.0	3294	ABX34502	Abx34502 Human mdd
19	390.6	43.1	2611	AAF21812	Aaf21812 Human bre
20	329.8	36.4	2412	ADR45468	Adh45468 Human mol

21	243.2	26.8	2469	12	ADH45478	Adh45478 Human mol
22	212.2	23.4	5282	12	ADQ64426	Adq64426 Novel hum
23	211.8	23.4	2044	4	AAI60392	Aai60392 Human pol
c 24	208.6	23.0	3828	13	ADR07582	Adr07582 Full leng
25	207.4	22.9	2979	4	ABLI17657	Abli17657 Drosophil
26	190.6	21.0	1958	4	AH18689	Aah18689 Human cDN
c 27	190.6	21.0	10468	4	AAK77961	Aak77961 Human imm
28	189.2	20.9	2217	10	ACA92443	ACA92443 DNA encod
29	189	20.9	1567	8	ACC46592	ACC46592 Human dit
30	174.2	19.2	2543	4	ABL17449	Abli17449 Drosophil
31	166	18.3	1728	5	AAS70158	Aas70158 DNA encod
32	164.4	18.1	2250	5	AAS69383	Aas69383 DNA encod
33	163.4	18.0	418	8	ABX52103	Abx52103 Bovine ES
34	144.8	16.0	4615	4	ABLI17448	Abli17448 Drosophil
c 35	136.8	15.1	5748	4	ABLI17656	Abli17656 Drosophil
36	131.8	14.5	1085	4	AAF63819	Aaf63819 Human sec
c 37	130	14.3	525	12	ACH77979	Ach77979 Human gen
38	117	12.9	492	9	ACH48041	Ach48041 Human lun
39	92.2	10.2	207	6	ABS69387	Abs69387 Novel mur
c 40	80	8.8	80	12	ADJ93402	Adj93402 Human BGS
41	79.4	8.8	1571	4	ABLI4515	Abli4515 Drosophil
42	71.8	7.9	996	10	ADC32299	Adc32299 Human nov
43	71.8	7.9	2259	12	ADQ64293	Adq64293 Novel hum
44	71.8	7.9	2417	10	ADC30440	Adc30440 Human nov
45	65.6	7.2	1579	8	ABX05228	Abx05228 Human nov

ALIGNMENTS

RESULT 1
ADJ93357
ID ADJ93357 standard; cDNA; 1838 BP.
XX AC ADJ93357;
XX DT 06-MAY-2004 (first entry)
XX DE Human BGS-42 cDNA sequence SeqID1.
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-gen; gastrointestinal-gen;
KW neuroprotective; endocrine-gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; neotropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.
Homo sapiens.
Key Location/Qualifiers
CDS 153..1778
/*tag= a
/product= "Human BGS-42 protein"
WO2004005487-A2.
15-JAN-2004.
09-JUL-2003; 2003WO-US021605.
09-JUL-2002; 2002US-0394725P.
(BRIM) BRISTOL-MYERS SQUIBB CO.
Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.
DR P-PSDB; ADJ93358.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
PS Claim 1; SEQ ID NO 1; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytotstatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.
XX
SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Query Match 100.0%; Score 906; DB 12; Length 1838;
Best Local Similarity 100.0%; Pred. No. 2.1e-207; Mismatches 0; Gaps 0;
Matches 906; Conservative 0; Indels 0;

QY 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 60
DB 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 428

QY 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCACTCCAAATTCAGAAAT 120
DB 429 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCACTCCAAATTCAGAAAT 488

QY 121 TACTTTTCCAGTGGCCAGGCTCTGCTGAATAGAAATCAGCTGTGAACCTCAGACGGAC 180
DB 489 TACTTTTCCAGTGGCCAGGCTCTGCTGAATAGAAATCAGCTGTGAACCTCAGACGGAC 548

QY 181 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCCGCAAGTCCCGGGCCGAGAC 240
DB 549 ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCCGCAAGTCCCGGGCCGAGAC 608

QY 241 ATAGTGTGCATGCACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCT 300
DB 609 ATAGTGTGCATGCACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCT 668

QY 301 TCCAGGGACAACAGTGGGTGTCAGAGATACATCGACGCCCGCTCATCTGTGAC 360
DB 669 TCCAGGGACAACAGTGGGTGTCAGAGATACATCGACGCCCGCTCATCTGTGAC 728

QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCAATCTGG 420
DB 729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCAATCTGG 788

QY 421 TTCTACAAGGAGAGTTACTTTGGGTTCTCAATCAGCGCTTCTCCCTGGACAAAGCTGGAC 480

DB 789 TTCTACAAGGAGAGTTACTTTGGGTTCTCAATCAGCGCTTCTCCCTGGACAAAGCTGGAC 848
QY 481 AGCGGCATCCACTCTGTGCAACAAACCGCTGCAGAAAGTACTCTGAAGAATGATGTGGGCCCGC 540
DB 849 AGCGGCATCCACTCTGTGCAACAAACCGCTGCAGAAAGTACTCTGAAGAATGATGTGGGCCCGC 908
QY 541 AGCCCCCTGCTGCTGCCCGCACACAACATGTGGACACAGGTTCCAGGAGTACTCTGCAG 600
DB 909 AGCCCCCTGCTGCTGCCCGCACACAACATGTGGACACAGGTTCCAGGAGTACTCTGCAG 968
QY 601 CGCAGAGGCGGTGGCGCGCTGTGGGGCAGCGTCACTACCCGTCATGTAAGAGGCCATC 660
DB 969 CGCAGAGGCGGTGGCGCGCTGTGGGGCAGCGTCACTACCCGTCATGTAAGAGGCCATC 1028
QY 661 GCCCAGCCCATCAAGGTGGCCCGCCAGCACACACGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 720
DB 1029 GCCCAGCCCATCAAGGTGGCCCGCCAGCACACACGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 1088
QY 721 TACGGGGCTGACTTCTGCTCTTGGGAGGAGCTTCAGGCCCTGCTGATCGAGATCAATTCC 780
DB 1089 TACGGGGCTGACTTCTGCTCTTGGGAGGAGCTTCAGGCCCTGCTGATCGAGATCAATTCC 1148
QY 781 AGCCCCCATCAAGGTGGCCCGCTGCGACCGCTCAGCGCCAGCTGTGTCACAGGTGCAGGAG 840
DB 1149 AGCCCCCATCAAGGTGGCCCGCTGCGACCGCTCAGCGCCAGCTGTGTCACAGGTGCAGGAG 1208
QY 841 GACACCATCAAGGTGGCCCGCTGCGACCGCTGCGACCGCTGCGACCGCTGCGACCGCTGCG 900
DB 1209 GACACCATCAAGGTGGCCCGCTGCGACCGCTGCGACCGCTGCGACCGCTGCGACCGCTGCG 1268
QY 901 AGGCAG 906
DB 1269 AGGCAG 1274

RESULT 2
ADJ93364
ID ADJ93364 standard; DNA; 3554 BP.
XX AC ADJ93364;
XX DT 06-MAY-2004 (first entry)
XX DE Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.
XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytotstatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; neurotropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1695..3320
FT /tag= a
FT /product= "Human tubulin tyrosine ligase protein"
XX
XX WO2004005487-A2.
XX 15-JAN-2004.
XX

PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
DR WPI; 2004-099381/10.
DR P-PSDB; ADJ93365.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
PS Example 4; SEQ ID NO 12; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
CC antiparkinsonian, antiarthritic, antiaesthatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the
CC tubulin tyrosine ligase protein consensus gene sequence which was used in
CC the exemplification of the invention.
XX
SQ Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;
Query Match 100.0%; Score 906; DB 12; Length 3554;
Best Local Similarity 100.0%; Pred. No. 2.5e-207;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGCCGCGAGTGGAGGAC 60
DB 1911 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGCCGCGAGTGGAGGAC 1970
QY 61 CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTCATCTCAATTCAGAAAT 120
DB 1971 CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTCATCTCAATTCAGAAAT 2030
QY 121 TACTTTTCGAGTGCACGCTGCTGCTGAATAGATCACTGCTGACCTCTGACCTCTGACGGAC 180
DB 2031 TACTTTTCGAGTGCACGCTGCTGCTGAATAGATCACTGCTGACCTCTGACGGAC 2090
QY 181 ATTGACGGGCTCCGGAACATCTGATATATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
DB 2091 ATTGACGGGCTCCGGAACATCTGATATATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 2150
QY 241 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTT 300
DB 2151 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTT 2210
QY 301 TCAGGGACAAAGTGGGTGGTCCAGAGTACATCGAGACGCCGCTGCTCATCTGTGAC 360

DB 2211 TCCAGGGGCAACAGTGGGTGGTCCAGAGTACATCGAGACGCCGCTCATCTGTGAC 2270
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTTGACCATCTGG 420
DB 2271 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTTGACCATCTGG 2330
QY 421 TTCTACAAGGAGAGTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
DB 2331 TTCTACAAGGAGAGTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 2390
QY 481 AGGCCCATCCACCTGTGCAACCAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGGCGCG 540
DB 2391 AGGCCCATCCACCTGTGCAACCAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGGCGCG 2450
QY 541 AGCCCCCTGTGCTCCCGCAACACATGTGAGCAGCAGCAACAGGTTCCAGGAGTACCTGTCAG 600
DB 2451 AGCCCCCTGTGCTCCCGCAACACATGTGAGCAGCAGCAACAGGTTCCAGGAGTACCTGTCAG 2510
QY 601 CGCCAGGGCGGTGGCGCGCTGTGGGCGAGGTCATCTACCGTCCATGAGAGAGGCCATC 660
DB 2511 CGCCAGGGCGGTGGCGCGCTGTGGGCGAGGTCATCTACCGTCCATGAGAGAGGCCATC 2570
QY 661 GCCCAGCCCATGAAGTGGCCAGGACCACTGGAGCTCGCAAGAACAGCTTTGAGCTC 720
DB 2571 GCCCAGCCCATGAAGTGGCCAGGACCACTGGAGCTCGCAAGAACAGCTTTGAGCTC 2630
QY 721 TAGGGGCTGACTTCTCTTGGAGGGACTTCAGGCCCTGGCTGATCGAGATCAATTC 780
DB 2631 TAGGGGCTGACTTCTCTTGGAGGGACTTCAGGCCCTGGCTGATCGAGATCAATTC 2690
QY 781 AGCCCCCACCATGACCCGCTCCAGCGGTCACGGCCCGCTGTCACAGGTGTCAGAGTGCAGGAG 840
DB 2691 AGCCCCCACCATGACCCGCTCCAGCGGTCACGGCCCGCTGTCACAGGTGTCAGAGGAG 2750
QY 841 GACACCATCAAGTGGCCGCTGGAGCCGCTGTGACATCGGCAACTTCAGAGCTCTGTGG 900
DB 2751 GACACCATCAAGTGGCCGCTGGAGCCGCTGTGACATCGGCAACTTCAGAGCTCTGTGG 2810
QY 901 AGGCAG 906
DB 2811 AGGCAG 2816
RESULT 3
ADJ93362
ID ADJ93362 standard; DNA; 1859 BP.
XX
AC ADJ93362;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone B SegID10.
XX
DE testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiaesthatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
PN WO2004005487-A2.
XX

PD 15-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Example 4; SEQ ID NO 10; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, respiratory
CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;

Query Match 89.1%; Score 806.8; DB 12; Length 1859;
Best Local Similarity 95.0%; Pred. No. 1.3e-183;
Matches 861; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 1 GAGGACATCGACACGTACAGATCCGCTGGAGGACCTCAGTCAGCCGAGTGGGAGGAC 60
DB 263 GAGGACATCGACACGTACAGATCCGCTGGAGGACCTCAGTCAGCCGAGTGGGAGGAC 322
QY 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT 120
DB 323 CTGACCCAGCAGTACTACTCCCTCGTTCA----- 351
QY 121 TACTTTTCAGTCGACGCTCTGTGAATAGAAATCACTGTGTGAACCTCAGACGGAC 180
DB 352 -----TGCCAGGCTCTGCTGAATAGAAATCACGCTCTGTGAACCTCAGACGGAC 399
QY 181 ATTGACGGCTCGGAAACATCTGGATTATTAAGCCCGCCGCAAGTCCCGGGCCGAGAC 240
DB 400 ATTGACGGCTCGGAAACATCTGGATTATTAAGCCCGCCGCAAGTCCCGGGCCGAGAC 459
QY 241 ATAGTGTGATGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTCAGACCAACCTCTT 300
DB 460 ATAGTGTGATGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTCAGACCAACCTCTT 519
QY 301 TCCAGGGACAACAAAGTGGGTGTCAGAAAGTATATCGAGACGCGCGTCTCATCTGTGAC 360

Db 520 TCCAGGGACAACAAGTGGGTGGTCCAGAAAGTATCATCGAGACCCCGTCTCATCTGTGAC 579
QY 361 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCACGACTCGGAACCCCGCTGACCATCTCG 420
Db 580 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCACGACTCGGAACCCCGCTGACCATCTCG 639
QY 421 TTCTACAAGGAGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480
Db 640 TTCTACAAGGAGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 699
QY 481 AGCGCATTCACCTGTGCAACCAACCGCTTCAGAAAGTACTCAAGATCATGTGGCCCG 540
Db 700 AGCGCATTCACCTGTGCAACCAACCGCTTCAGAAAGTACTCAAGATCATGTGGCCCG 759
QY 541 AGCCCGCTTCTGCTCCCGCACACAAATGTGGACACGACGAGTTCAGGAGTACCTGTCAG 600
Db 760 AGCCCGCTTCTGCTCCCGCACACAAATGTGGACACGAGTTCAGGAGTACCTGTCAG 819
QY 601 CGCCAGGGCGGTGGCGCGCTGTGGGGCAGCGTCATCTACCCCTCCATGAAGAAGCCATC 660
Db 820 CGCCAGGGCGGTGGCGCGCTGTGGGGCAGCGTCATCTACCCCTCCATGAAGAAGCCATC 879
QY 661 GCCCAGCCATGAAGTGGCCCGCAGGACCACTGGAGCCCTCGCAAGAACAGCTTGGAGCTC 720
Db 880 GCCCAGCCATGAAGTGGCCCGCAGGACCACTGGAGCCCTCGCAAGAACAGCTTGGAGCTC 939
QY 721 TAGCGGGCTGACTTCTGCTCTCGGAGGAGTTCAGGCGCTGCTGATCGAGATCAATTC 780
Db 940 TAGCGGGCTGACTTCTGCTCTCGGAGGAGTTCAGGCGCTGCTGATCGAGATCAATTC 999
QY 781 AGCCCCCACCATGACCCCGTCCACGCGGTTCACGCGCCAGCTGTGTCACAGTGCAGGAG 840
Db 1000 AGCCCCCACCATGACCCCGTCCACGCGGTTCACGCGCCAGCTGTGTCACAGTGCAGGAG 1059
QY 841 GACACATCAAGTGGCGCTGACCGCAGCTGTGATCGGCAACTTCAGCTCTCTGTGG 900
Db 1060 GACACATCAAGTGGCGCTGACCGCAGCTGTGATCGGCAACTTCAGCTCTCTGTGG 1119
QY 901 AGGCGAG 906
Db 1120 AGCCCG 1125

RESULT 4
ADJ93363
ID ADJ93363 standard; DNA; 3465 BP.
XX
AC ADJ93363;
XX
DT 06-MAY-2004 (first entry)
XX
XX Human BGS-42 protein-related DNA clone C SeqID11.
DE
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; neurotropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
XX WO2004/005487-A2.
XX

PD 15-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
DR WPI; 2004-099381/10.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
PS Example 4; SEQ ID NO 11; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytosolic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match 84.1%; Score 762; DB 12; Length 3465;
Best Local Similarity 92.6%; Pred. No. 9.1e-173;
Matches 839; Conservative 0; Mismatches 0; Indels 67; Gaps 1;

QY 1 GAGGACATCGACAGCTCAGCAGATGCGGTGGAGGACCTCACTGAGGCGGAGTGGAGGAC 60
DB 1894 GAGGACATCGACAGCTCAGCAGATGCGGTGGAGGACCTCACTGAGGCGGAGTGGAGGAC 1953
QY 61 CTGACCCAGCAGTACTACTCCCTGGTTCATGGCGATCTTTCATCTCCAAATCAAGAAAT 120
DB 1954 CTGACCCAGCAGTACTACTCCCTGGTTCATGGCGATCTTTCATCTCCAAATCAAGAAAT 2013
QY 121 TACTTTTCGAGTCCGAGGCTCTGCTGAATAGATCACTGTGTAACCTCAGACGGAC 180
DB 2014 TACTTTTCGAGTCCGAGGCTCTGCTGAATAGATCACTGTGTAACCTCAGACGGAC 2073
QY 181 ATTGACGGGTCCGGAACATCTGGATTATAAAGCCGCGGCAAGTCCCGGGGCCGAGAC 240
DB 2074 ATTGACGGGTCCGGAACATCTGGATTATAAAGCCGCGGCAAGTCCCGGGGCCG--- 2129
QY 241 ATAGTGTGTCATGGACCGCTGTGGAGGAGATCTCTGGAGTGGCAGCTGCAGACCCCTCTT 300
DB 2130 ----- 2129
QY 301 TCAGGAGCAACAAGTGGGTGGTCCAGAGTACATCGAGAGCGCGGTGCTCATCTGTGAC 360

DB 2130 ---AGGACAAAGTGGGTGGTCCAGAGTACATCGAGAGCGCGTCTCATCTGTGAC 2186
QY 361 ACCAAGTTCGACATCAGACAGATGTTCTCTGTTCACGGAACCTCCCTGACCATCTGG 420
DB 2187 ACCAAGTTCGACATCAGACAGATGTTCTCTGTTCACGGAACCTCCCTGACCATCTGG 2246
QY 421 TTCTACAGGAGGATTTACTTGGCGTTCTCAACTCAGGCGTCTCCCTGGCAAGCTGGAC 480
DB 2247 TTCTACAGGAGGATTTACTTGGCGTTCTCAACTCAGGCGTCTCCCTGGCAAGCTGGAC 2306
QY 481 AGGCGCATCCACCTGTGCAACACGCGTCCAGAGTACCTGAAGATATGATGGGCGCG 540
DB 2307 AGGCGCATCCACCTGTGCAACACGCGTCCAGAGTACCTGAAGATATGATGGGCGCG 2366
QY 541 AGCCCGCTGTGCTCCCGCACCAACATGTGTGACACAGCAACAGGTTCCAGGAGTACCTGCAG 600
DB 2367 AGCCCGCTGTGCTCCCGCACCAACATGTGTGACACAGCAACAGGTTCCAGGAGTACCTGCAG 2426
QY 601 CGCAGGCGCGTGGCGCGCTGTGGGCGAGCGTCTATCCCGTCCATGAGAGGCGCATC 660
DB 2427 CGCAGGCGCGTGGCGCGCTGTGGGCGAGCGTCTATCCCGTCCATGAGAGGCGCATC 2486
QY 661 GCCCAGCCCATGAAGTGGCGCCAGGACCACTGGAGCCTCGCAAGAACACAGCTTTGAGCTC 720
DB 2487 GCCCAGCCCATGAAGTGGCGCCAGGACCACTGGAGCCTCGCAAGAACACAGCTTTGAGCTC 2546
QY 721 TAGCGGCGTGAATTCCTTCTGTGGAGGAGCTTCAGGCGCTGATCGAGATCAATTC 780
DB 2547 TAGCGGCGTGAATTCCTTCTGTGGAGGAGCTTCAGGCGCTGATCGAGATCAATTC 2606
QY 781 AGCCCGCATGACCGCTGTCAGCGCGTTCAGCGCGGCTGTCAGAGTGCAGGAG 840
DB 2607 AGCCCGCATGACCGCTGTCAGCGCGTTCAGCGCGGCTGTCAGAGTGCAGGAG 2666
QY 841 GACACCATCAAGTGGCGCGTGGAGCCGAGCTGACATCGGCAACTTCGAGCTCTCTGG 900
DB 2667 GACACCATCAAGTGGCGCGTGGAGCCGAGCTGACATCGGCAACTTCGAGCTCTCTGG 2726
QY 901 AGGCAG 906
DB 2727 AGGCAG 2732

RESULT 5
ADJ93361
ID ADJ93361 standard; DNA; 1939 BP.
XX
AC ADJ93361;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone A SeqID9.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiaesthetic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
PN W02004005487-A2.
XX

PD 15-JAN-2004.
 XX 09-JUL-2003; 2003WO-US021605.
 XX 09-JUL-2002; 2002US-0394725P.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Feder JN, Wu S, Nelson TC;
 XX WPI; 2004-099381/10.
 XX
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 XX useful for preventing, treating or ameliorating a medical condition, e.g.
 XX aberrant cellular proliferation, reproductive disorders or testicular
 XX disorders.
 XX
 XX Example 4; SEQ ID NO 9; 343pp; English.
 XX
 XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
 XX be useful for the development of compounds with a cytotstatic, respiratory
 XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 XX antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 XX antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 XX immunosuppressive, antiseborrheic or dermatological activity acting as
 XX tyrosine ligase modulators. In addition, the disclosed sequences may be
 XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 XX used for diagnosing a pathological condition or a susceptibility to a
 XX pathological condition in a subject, and for preventing, treating or
 XX ameliorating a medical condition, such as a disorder related to aberrant
 XX tubulin ligase activity, a disorder related to aberrant tubulin-
 XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
 XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
 XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 XX neural disorders, brain cancer, liver cancer, or proliferative condition
 XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 XX polypeptide, polynucleotide, or their modulators are also useful for
 XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 XX -42 polypeptide can be used as a preventive agent for immunological
 XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 XX disease or scleroderma. The antibodies may be used to purify, detect and
 XX target the BGS-42 polypeptides. The present sequence is that of a DNA
 XX clone sequence which is related to the invention.
 XX
 XX Sequence 1939 BP; 421 A; 592 C; 589 G; 337 T; 0 U; 0 Other;
 XX
 XX Query Match 83.9%; Score 760.4; DB 12; Length 1939;
 XX Best Local Similarity 92.5%; Pred. No. 1.9e-172;
 XX Matches 838; Conservative 0; Mismatches 1; Indels 67; Gaps 1;
 XX
 XX 1 GAGGACATCGACAGTCTACGAGTCCGCTGGAGGACCTCACTGAGGCCGAGTGGGAGGAC 60
 XX |||||
 XX 369 GAGGACATCGACAGTCTACGAGTCCGCTGGAGGACCTCACTGAGGCCGAGTGGGAGGAC 428
 XX |||||
 XX 61 CTGACCCAGCAGTACTCTCTCTGTTCTATGCGATGCTTTCTATCTCCAAATCAAGAAAT 120
 XX |||||
 XX 429 CTGACCCAGCAGTACTCTCTCTGTTCTATGCGATGCTTTCTATCTCCAAATCAAGAAAT 488
 XX |||||
 XX 121 TACTTTTCCAGTGCAGGCTCTGCTGAATAGAAATACGCTGTGTGAACCTTCAGACGGAC 180
 XX |||||
 XX 489 TACTTTTCCAGTGCAGGCTCTGCTGAATAGAAATACGCTGTGTGAACCTTCAGACGGAC 548
 XX |||||
 XX 181 ATTGACGGGCTCCGAGCATCTGGATTATTAAGCCCGGGCCAGTCCCGGGGGCGAGAC 240
 XX |||||
 XX 549 ATTGACGGGCTCCGAGCATCTGGATTATTAAGCCCGGGCCAGTCCCGGGGGCGG---- 604
 XX |||||
 XX 241 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGTGCAGACCAACCTCTTT 300
 XX |||||
 XX 605 ----- 604
 XX 301 TCCAGGGACAAAGTGGGTGTGCCAAGAGTACATCGAGACGCCGCTGCTCATCTGTGAC 360
 XX |||||

Db 605 ---AGGGAACAACAGTGGGTGGTCCAGAGTAGTACATCGAGACGCCGCTGCTCATCTGTGAC 661
 QY 361 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCATCTGTG 420
 Db 662 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCATCTGTG 721
 QY 421 TTCTACAAGGAGAGTGTACTTTCGCGTTTCACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480
 Db 722 TTCTACAAGGAGAGTGTACTTTCGCGTTTCACTCAGCGCTTCTCCCTGGACAAGCTGGAC 781
 QY 481 AGCGCATTCACCTGTGCAACAAACCGCTTCAGAGTACCTGAAAGATGATGTGGCCCGC 540
 Db 782 AGCGCATTCACCTGTGCAACAAACCGCTTCAGAGTACCTGAAAGATGATGTGGCCCGC 841
 QY 541 AGCCCCCTGCTCCCGCGCACACAAACATGTGGACACGACAGGTTCAGGAGTACCTGCGAC 600
 Db 842 AGCCCCCTGCTCCCGCGCACACAAACATGTGGACACGACAGGTTCAGGAGTACCTGCGAC 901
 QY 601 CGCCAGGCGCGTGGCGCGCTGTGGGGCAGCGTCATCTACCCGCTCCATGAAGAGGCCATC 660
 Db 902 CGCCAGGCGCGTGGCGCGCTGTGGGGCAGCGTCATCTACCCGCTCCATGAAGAGGCCATC 961
 QY 661 GCCCAGCGCATGAAGTGGGCCCAGGACACACGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 720
 Db 962 GCCCAGCGCATGAAGTGGGCCCAGGACACACGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 1021
 QY 721 TACGGGGCTGACTTCGCTCTGGGAGGAGTTCAGGCGCTGCTGATCGAGATCAATTC 780
 Db 1022 TACGGGGCTGACTTCGCTCTGGGAGGAGTTCAGGCGCTGCTGATCGAGATCAATTC 1081
 QY 781 AGCCCCACCATGCACCCCGCTCCAGCGCGGTTCAGGCGCGCTGTGTCACAGTGCAGGAG 840
 Db 1082 AGCCCCACCATGCACCCCGCTCCAGCGCGGTTCAGGCGCGCTGTGTCACAGTGCAGGAG 1141
 QY 841 GACACCATCAAGTGGCGCGTGGACCGCGAGCTGTGATCGGCAACTTCGAGCTCTCTGTGG 900
 Db 1142 GACACCATCAAGTGGCGCGTGGACCGCGAGCTGTGATCGGCAACTTCGAGCTCTCTGTGG 1201
 QY 901 AGGCGAG 906
 Db 1202 AGGCGAG 1207
 XX
 XX RESULT 6
 XX ADJ93359
 XX ID ADJ93359 standard; cDNA; 726 BP.
 XX AC ADJ93359;
 XX DT 06-MAY-2004 (first entry)
 XX DE Human BGS-42 cDNA sequence SeqID3.
 XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 XX BGS-42 polypeptide; cytotstatic; respiratory-Gen; gastrointestinal-Gen;
 XX neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 XX osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 XX anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 XX dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 XX tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 XX testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 XX gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 XX brain cancer; liver cancer; proliferative condition; testis; lung;
 XX small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 XX emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 XX Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 XX sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX CDS 1..726

/*tag= a
/product= "Human BGS-42 protein"
/partial
/notes= "No start or stop codon"

WO2004005487-A2.

15-JAN-2004.

09-JUL-2003; 2003WO-US021605.

09-JUL-2002; 2002US-0394725P.

(BRIM) BRISTOL-MYERS SQUIBB CO.

Feder JN, Wu S, Nelson TC;

WPI; 2004-099381/10.

New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, useful for preventing, treating or ameliorating a medical condition, e.g. aberrant cellular proliferation, reproductive disorders or testicular disorders.

Disclosure; SEQ ID NO 3; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase-like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cytostatic, respiratory -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, antiinflammatory, anabolic, hypertensive, osteopathic, nootropic, antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial, immunosuppressive, anticancer or dermatological activity acting as tyrosine ligase modulators. In addition, the disclosed sequences may be useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a partial cDNA which encodes the human BGS-42 protein of the invention.

SQ Sequence 726 BP; 157 A; 227 C; 214 G; 128 T; 0 U; 0 Other;

Query Match 80.1%; Score 726; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 2.6e-164;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 ATTGACGGGCTCCGGAACATCTGATTATTAAGCCGCGGCCCAAGTCCCGGGCCGAGAC 240
DB 1 ATTGACGGGCTCCGGAACATCTGATTATTAAGCCGCGGCCCAAGTCCCGGGCCGAGAC 60
QY 241 ATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGTGGCAGTGCGACACCCCTCTT 300
DB 61 ATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGTGGCAGTGCGACACCCCTCTT 120
QY 301 TCAGGAGAACAAAGTGGTGGTCCAGAACTATCATGAGAGCGCGTGTCTATCTGTGAC 360
DB 121 TCAGGAGAACAAAGTGGTGGTCCAGAACTATCATGAGAGCGCGTGTCTATCTGTGAC 180
QY 361 ACCAAGTTCACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCATCTGG 420

Db 181 ACCAAGTTCACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCATCTGG 240
QY 421 TTCTACAGGAGAGTACTTTCGGTTCCTAACTCAGCGCTTCTCCCTGGGACAGCTGGAC 480
Db 241 TTCTACAGGAGAGTACTTTCGGTTCCTAACTCAGCGCTTCTCCCTGGGACAGCTGGAC 300
QY 481 AGCGCCATCCACCTGTGCAACAAACGCGTCCAGAGTACCTGAAGATATGTGGCCGC 540
Db 301 AGCGCCATCCACCTGTGCAACAAACGCGTCCAGAGTACCTGAAGATATGTGGCCGC 360
QY 541 AGCCCCCTGTGTCGCCGACACAAACATGTGTGACCAAGTTCAGGAGTACCTGAC 600
Db 361 AGCCCCCTGTGTCGCCGACACAAACATGTGTGACCAAGTTCAGGAGTACCTGAC 420
QY 601 CGCAGGCGCGTGGCGCGTGTGGGAGCGTGTATACCCGTCATGAAGAGGCCATC 660
Db 421 CGCAGGCGCGTGGCGCGTGTGGGAGCGTGTATACCCGTCATGAAGAGGCCATC 480
QY 661 GCCCAGCCATGAAGTGGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTC 720
Db 481 GCCCAGCCATGAAGTGGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTC 540
QY 721 TAGCGGCGTGCATCTCGTTCCTTGGGAGGAGCTTCAGGCGCTGGCTGATCGAGATCAATTC 780
Db 541 TAGCGGCGTGCATCTCGTTCCTTGGGAGGAGCTTCAGGCGCTGGCTGATCGAGATCAATTC 600
QY 781 AGCCCCCAGCATGACCCGCTCCAGCGGTCCAGCGGTCCAGCGGTGTCGACAGGTGACAGAG 840
Db 601 AGCCCCCAGCATGACCCGCTCCAGCGGTCCAGCGGTCCAGCGGTGTCGACAGGTGACAGAG 660
QY 841 GACACCATCAAGTGGCGCGTGGAGCCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 900
Db 661 GACACCATCAAGTGGCGCGTGGAGCCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 720
QY 901 AGGCAG 906
Db 721 AGGCAG 726

RESULT 7
ACH91699/c
ID ACH91699 standard; DNA; 490 BP.
XX ACH91699;
XX 29-JUL-2004 (first entry)
XX Human genome derived single exon probe #24894.
XX Human; probe; ss; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX Homo sapiens.
XX US2003194704-A1.
XX 16-OCT-2003.
XX 03-APR-2002; 2002US-00029386.
XX 03-APR-2002; 2002US-00029386.
XX (PENN/) PENN S G.
XX (RANK/) RANK D R.
XX (HANZ/) HANZEL D K.
XX Penn SG, Rank DR, Hanzel DK;
XX WPI; 2004-119264/12.
XX New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for

PT surveying tissues.
XX Claim 1; SEQ ID NO 24894; 80pp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ Sequence 490 BP; 77 A; 150 C; 168 G; 95 T; 0 U; 0 Other;

Query Match 47.1%; Score 427; DB 12; Length 490;
Best Local Similarity 100.0%; Pred. No. 1.4e-92;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 CAGCGCCATCCACCTGTGCAACACGCCGTCCAGAGTACCTGAAAGATGATGTGGCGG 539
DB |||||||
QY 427 CAGCGCCATCCACCTGTGCAACACGCCGTCCAGAGTACCTGAAAGATGATGTGGCGG 368
DB |||||||
QY 540 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 599
DB |||||||
QY 367 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 308
DB |||||||
QY 600 GCGCCACGCGCGTGGCGCGGTGTGGGCGACGCTATCTACCGTCCATGAAGGCCAT 659
DB |||||||
QY 307 GCGCCACGCGCGTGGCGCGGTGTGGGCGACGCTATCTACCGTCCATGAAGGCCAT 248
DB |||||||
QY 660 GCGCCACGCGCATGAGGTGGCGCGCGCACGACGCTGGAGCTCCGACAAACAGCTTCAGCT 719
DB |||||||
QY 247 GCGCCACGCGCATGAGGTGGCGCGCGCACGACGCTGGAGCTCCGACAAACAGCTTCAGCT 188
DB |||||||
QY 720 CTACGGGGGTGACTTCGTCTCTGGGAGGACTTCAGGCCCTGTGCTGATCGAGATCAATTC 779
DB |||||||
QY 187 CTACGGGGGTGACTTCGTCTCTGGGAGGACTTCAGGCCCTGTGCTGATCGAGATCAATTC 128
DB |||||||
QY 780 CAGCCCCACATGCAACCCGCTCAACGCCGCTCAAGGCCAGCTGTGTGCAACAGTTCAGGA 839
DB |||||||
QY 127 CAGCCCCACATGCAACCCGCTCAACGCCGCTCAAGGCCAGCTGTGTGCAACAGTTCAGGA 68
DB |||||||
QY 840 GGACACATCAAGGTGGCGCGTGAACGCGAGCTGTGACATCGGCAACTTCAGCTCCTGTG 899
DB |||||||
QY 67 GGACACATCAAGGTGGCGCGTGAACGCGAGCTGTGACATCGGCAACTTCAGCTCCTGTG 8

QY 900 GAGGCAG 906
DB |||||||
DB 7 GAGGCAG 1

RESULT 8
ADQ17814/C
ID ADQ17814 standard; DNA; 101270 BP.
XX
AC ADQ17814;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 631.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 631; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytotstatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;

Query Match 47.1%; Score 427; DB 12; Length 101270;
Best Local Similarity 100.0%; Pred. No. 6.1e-92;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 CAGCGCCATCCACCTGTGCAACACGCCGTCCAGAGTACCTGAAAGATGATGTGGCGG 539
DB |||||||
DB 44821 CAGCGCCATCCACCTGTGCAACACGCCGTCCAGAGTACCTGAAAGATGATGTGGCGG 44762
QY 540 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 599
DB |||||||
DB 44761 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 44702
QY 600 GCGCCACGCGCGTGGCGCGGTGTGGGCGAGCTCTACTCCGTCCTCATGAAGGCCAT 659
DB |||||||
DB 44701 GCGCCACGCGCGTGGCGCGGTGTGGGCGAGCTCTACTCCGTCCTCATGAAGGCCAT 44642

QY 660 CGCCACGCGATGAAGTGGCCAGGACCACTGGAGCTCGCAAGACAGCTTTGAGCT 719
Db |||||
44641 CGCCACGCGATGAAGTGGCCAGGACCACTGGAGCTCGCAAGACAGCTTTGAGCT 44582
QY 720 CTACGGGGCTGACTTCCTCTTGGAGGGGACTTCAGGCGCTGGCTGATCGAGATCAATTC 779
Db |||||
44581 CTACGGGGCTGACTTCCTCTTGGAGGGGACTTCAGGCGCTGGCTGATCGAGATCAATTC 44522
QY 780 CAGCCCCACCATGACCCGCTCCACCGCGGTACCGCCGATGTCAGAGTGCAGGAGCA 839
Db |||||
44521 CAGCCCCACCATGACCCGCTCCACCGCGGTACCGCCGATGTCAGAGTGCAGGAGCA 44462
QY 840 GGACACCATCAAGTGGCGCTGGACCGCAGCTGTCAGATCGGCAATTCGAGTCCCTGTG 899
Db |||||
44461 GGACACCATCAAGTGGCGCTGGACCGCAGCTGTCAGATCGGCAATTCGAGTCCCTGTG 44402
QY 900 GAGGCAG 906
Db |||||
44401 GAGGCAG 44395

RESULT 9
AAC77214
ID AAC77214 standard; cDNA; 2881 BP.
XX
AC AAC77214;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2769 polynucleotide sequence SEQ ID NO:5537.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
PI
XX WPI; 2000-602362/57.
XX
XX P-PSDB; AAB43005.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 5; Page 4718-4720; 5507pp; English.
XX
XX

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX Sequence 2881 BP; 640 A; 812 C; 840 G; 589 T; 0 U; 0 Other;
Query Match 44.2%; Score 400.4; DB 3; Length 2881;
Best Local Similarity 66.1%; Pred. No. 5.5e-86;
Matches 578; Conservative 0; Mismatches 296; Indels 0; Gaps 0;
QY 3 GGACATCGACACGTCAGCAGATCCGCTGGAGGACCTCACTGAGCGCCGAGTGGAGGACCT 62
Db |||||
894 GGACATCGACACGACCTGGAGGCGCCGCTGTACTCACCCCGAGGGCTGGTCCCTCTT 953
QY 63 GACCCAGAGTACTACTCCCTCTGTCATGCGGATGCTTTTCATCTCCAAATTCAGAAATTA 122
Db |||||
954 CCTCCAGCGCTACTACCAAGTGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCA 1013
QY 123 CTTTTCGAGTCGACGCTCTGCTGAATAGAAATCAGTCTGTAACCTCAGACGGACAT 182
Db |||||
1014 GTTCCAGCGCTGTGAGGACATCTCGACGAGCTGAGCGCGCTGTATCCAGATAGACAT 1073
QY 183 TGACGGGCTCCGGAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGCCGAGACAT 242
Db |||||
1074 GGAAGGGGATCGCAACATCTGGATCTGTAAGCCAGGACCAAGTCCCGTGGACGAGCAT 1133
QY 243 AGTGTGATGAGACCGTGTGGAGAGATCTTGGAGCTGGGAGCTGCAGACCACTCTTTC 302
Db |||||
1134 CATGTGATGAGACCACTGTGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGTGAT 1193
QY 303 CAGGGACAACAGTGGTGTCTCCAGAGTACATCGAGACGCGCTCATCTGTGACAC 362
Db |||||
1194 GAAGGACGGCAAGTGGTGTGAGAGTATATTTGAGCGCGCCCTCTCATCTTTGGCAC 1253
QY 363 CAAAGTTTCAGACATCAGACAGTGGTTCTCTCGTACGGACTTGAACCCCTGACCATCTGGTT 422
Db |||||
1254 CAAAGTTTCAGACATCAGACAGTGGTTCTCTCGTAACTGACTGGAACCCACTTACCGTGTGTT 1313
QY 423 CTACAAGGAGAGTACTTGGCGTTCTCAACTAGCGCTTCTCCCTGGACAAAGCTGGACAG 482
Db |||||
1314 CTACCGCGACAGTATATCGCTTTTCCAGCAGCGCTTCTCCCTGGAAGAACCTGGACAA 1373
QY 483 CGCCATCCACCTGTGCAACAACCGCTCCAGAGTACCTGGAAGTATGATGTGGGCGCAG 542
Db |||||
1374 CTCAAGTGCACCTGTGCAACAACCTCCATCCAGAGACCTTGGAGAACTCATGTCATCGCGCA 1433
QY 543 CCCCCCTGCTCCCGCACACAACATGTGGACCAACAGAGTTCCAGGAGTACCTGACGCG 602
Db |||||
1434 TCCACTGCTTCCGCGACAGACATGTGTGTAGCCAGAGGTTCCAGGCCCACTTGCAGGA 1493
QY 603 CCAGGCGCGTGGCGCGCTGTGGGCGAGCGTCACTATCCCGTCCATGAAGAGCCATCGC 662
Db |||||
1494 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGTCCTGGCATGAAGGATGCTGTGAT 1553
QY 663 CCAGCCCATGAAGTGGCGCCCGACGACCATCGTGGAGCTCCGCAAGAACAGCTTTGAGCTTA 722
Db |||||

Db 1554 CCACGCACTTCAGACCTCCAGACACCCGTGCAGTGTGGAAGCCAGCTTTGAGCTCTA 1613

QY 723 CGGGGTGACTTCGTCTCTTGGGAGGACTTCAGGCCCTCGCTGATCGAGATCAATTCACG 782

Db 1614 TGGCGTGACTTCGTCTGGGGAGGACTTCAGGCCCTCGCTGATCGAGATCAATTCACG 1673

QY 783 CCCACATGACACCGTGCACCGCGGTGCAGGCCAGCTGTGTGACAGGTGAGGAGGA 842

Db 1674 CCCACGATGGACACCTCCACAGCAGTCACTGCCCGCTCTGTGCTGGGTGCAAGCTGA 1733

QY 843 CACCATCAAGTGGCGTGGACCGCAGCTGTGAC 876

Db 1734 CACCCTGCGGTGGTCAATTGACCGGAGGTGGAC 1767

RESULT 10

AAH16735

ID AAH16735 standard; cDNA; 2326 BP.

XX AC AAH16735;

XX AC

XX XX

XX XX

XX 26-JUN-2001 (first entry)

XX XX

DE Human cDNA sequence SEQ ID NO:15920.

XX XX

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX XX

XX Homo sapiens.

XX XX

XX EP1074617-A2.

XX XX

XX 07-FEB-2001.

XX XX

XX 28-JUL-2000; 2000EP-00116126.

XX XX

XX 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX XX

XX (HELI-) HELIX RES INST.

XX XX

XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR XX

XX XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

PT PT

PT PT

XX XX

XX Claim 8; SEQ ID NO 15920; 2537pp + Sequence Listing; English.

PS XX

XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent CC oligonucleotides, all of which are used in the exemplification of the CC present invention

XX CC

SQ Sequence 2326 BP; 524 A; 701 C; 642 G; 459 T; 0 U; 0 Other;

Query Match 44.0%; Score 398.8; DB 4; Length 2326;

Best Local Similarity 66.3%; Pred. No. 1.3e-85;

Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 3 GGACATCGACAGTTCAGCAGATGCCGTGAGAGACCTCACTGAGCCGAGTGGGAGACCT 62

Db 732 GGACATCGACAGGACCTGGAGGCCCTGCTGATCACTCACCCCGAGGGTGGTCCCTCT 791

QY 63 GACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAATTA 122

Db 792 CCTCCAGCGCTACTACCAAGTGTCCAGAGGGGAGAACTCAGGCACCTCGACACTCA 851

QY 123 CTTTTCGCAAGTCCAGGCTCTCTGTAATAGAAATCACTGTGAACCTCAGACGACAT 182

Db 852 GGTCCAGCGCTGTGAGGACATCTGCAGCAGCTGCAGGCCGTGGTACCCAGATAGACAT 911

QY 183 TCACGGGCTCCGGAACATCTGGAATTATAAGCCCGCGGCCAAGTCCCGGGCCGACAT 242

Db 912 GGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAGTCCCGCGGACGAGCAT 971

QY 243 AGTGTGCATGGACCGTGTGGAGGAGATCTCGAGCTGCGAGCTGCAGACACACCTCTTTC 302

Db 972 CATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGTGAACGGCAACCCCGTGTGAT 1031

QY 303 CAGGGACAAACAGTGGTGGTCCAGAAATACATCGAGACCGCGCTGCTCATCTGTGACAC 362

Db 1032 GAAGGACGCAAGTGGTGGTGCAGAAATATATTGAGCGGCCCTCTCATCTTTGGCAC 1091

QY 363 CAAAGTTCGACATCAGACAGTGGTTCCTGTCAGGACTGGAACCCCTGACCATCTGGTT 422

Db 1092 CAAAGTTTGAACCTCAGACAGTGGTTCCTGGTAACTGACTGGAACCCCACTTACCGTGGTT 1151

QY 423 CTACAAGGAGAGTTACTTCGCGGTTCTCAACTCAGGCGTTCTCCCTGGACAAGCTGGACAG 482

Db 1152 CTACCGCGACAGCTATATCCGCTTTTCACGCGAGCCCTTCTCCCTGAAGAACCTGACAA 1211

QY 483 CGCCATCCACTGTGCAACACCGCGTCCAGAAATACCTGAAAGATGATGTGGCCCGCAG 542

Db 1212 CTCAGTGCACCTGTGCAACAACTCCATCCAGAAGCACTGGAGAACTCATGTCATCGGCA 1271

QY 543 CCCCCTGCTGCCCGCACCAACATGTGGACCAAGTTCAGGAGTTCAGGAGTACTGTCAGCG 602

Db 1272 TCCACTGCTTCCGCCAGACAACATGTGGTCTAGCCAGAGGTTCCAGGCCACCTCGCAGGA 1331

QY 603 CCAGGCGCGTGGCGCGTGTGGGCGAGCGTCTCATACCGTCCATGAAGAAGGCCATCGC 662

Db 1332 GATGGTGCCTCCAAATGCTTGGTTCACCATCATGCTGCTGGCATGAAGATGCTGTGAT 1391

QY 663 CCACGCGATGAAGTGGCCGAGGACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTA 722

Db 1392 CCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTA 1451

QY 723 CGGGGCTGACTTCGTCTTGGGAGGAGCACTTCAGGCCCTGGCTGATCGAGATCAATTCAG 782

Db 1452 TGGCGCTGACTTTCGTGTTCGGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAACGCCAG 1511

QY 783 CCCCACCATGACCCGTCACGCCCGTCCAGCGCTCAGCGCTGTGTGACAGGTGTGAGGAGGA 842

Db 1512 CCCCAGATGGACCCCTCCAGAGAGTCACTGCTGCCCGCTCTGTGCTGGCGTGCAGCTGA 1571

QY 843 CACCATCAAGTGGCGGTGGACCCGCA 868

Db 1572 CACCCTGCGCGTGGTCAATTGACCGGA 1597

RESULT 11

AAS99894	
XX	ID AAS99894 standard; cDNA; 2380 BP.
XX	AC AAS99894;
XX	DT 12-MAR-2002 (first entry)
XX	DE Polynucleotide encoding human cytoskeleton-associated protein (CYSKP) #5.
XX	KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;
KW	cell proliferative disorder; inflammatory disorder; prion disease;
KW	vesicle trafficking disorder; gastrointestinal disorder; muscle disease;
KW	neurological disorder; cell motility disorder; reproductive disorder;
KW	spinal cord disease; central nervous system disorder; mental disorder;
KW	gene therapy; cancer.
XX	OS Homo sapiens.
XX	PN WO200185942-A2.
XX	PD 15-NOV-2001.
XX	PF 03-MAY-2001; 2001WO-US014355.
XX	PR 05-MAY-2000; 2000US-0201960P.
PR	08-MAY-2000; 2000US-0202729P.
PR	03-JUN-2000; 2000US-0209705P.
PR	07-JUN-2000; 2000US-0210149P.
PR	21-JUN-2000; 2000US-0213215P.
XX	PA (INCY-) INCYTE GENOMICS INC.
XX	PI Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;
PI	Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;
PI	Policky JL;
XX	WP1: 2002-062248/08.
DR	P-PSDB; AAU74334.
XX	XX
PT	New cytoskeleton-associated proteins and polynucleotides, useful for
PT	diagnosing, preventing and treating cell proliferative, autoimmune,
PT	inflammatory, neurological, cell motility, reproductive and muscle
PT	disorders.
XX	XX
PS	Claim 5; Page 171; 194pp; English.
XX	XX
CC	The invention relates to human cytoskeleton-associated polypeptides
CC	(CYSKP) and their associated polynucleotide sequences. The sequences are
CC	useful in the treatment of disorders associated with overexpression or
CC	underexpression of CYSKP in a patient. The disorders include cell
CC	proliferative disorders (such as cancer, actinic keratosis,
CC	arteriosclerosis, cirrhosis, hepatitis and psoriasis),
CC	autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,
CC	osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus
CC	and anaemia), vesicle trafficking disorders (such as
CC	hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),
CC	gastrointestinal disorders, prion diseases, neurological disorders (such
CC	as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,
CC	Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis
CC	and other motor neuron disorders), cell motility disorders, reproductive
CC	disorders (such as endometriosis and polycystic ovary syndrome), muscle
CC	disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,
CC	myocardial infarction, epilepsy and muscular dystrophy), spinal cord
CC	diseases, central nervous system disorders (such as Down syndrome and
CC	cerebral palsy) and mental disorders (such as anxiety and schizophrenia).
CC	Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP
CC	of the invention
XX	XX
SQ	Sequence 2380 BP; 580 A; 685 C; 670 G; 445 T; 0 U; 0 Other;
	Query Match 44.0%; Score 398.8; DB 6; Length 2380;
	Best Local Similarity 66.3%; Pred. No. 1.3e-85;
	Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY	3	GGACATCGACACGTACAGATCGCGTGGAGGACCTCACTGAGCCGAGTGGAGGACCT	62
DB	473	GGACATCGACACAGGACCTGGAGGCCCGCTGTACCTACCCCGAGGGCTGGTCCCTCTT	532
QY	63	GACCCAGCAGTACTACTCCCTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATTA	122
DB	533	CCTCCAGCGCTACTACCAAGTGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCA	592
QY	123	CTTTTCGAGTCCAGGCTCTGCTGAATAGAATCAAGTCTGTGAACCTCAGACGGACAT	182
DB	593	GGTCCAGCGCTGTGAGGACATCTCGACAGCTGCAGGCGGTGTATCCCCAGATAGACAT	652
QY	183	TGAGGGGCTCCGAAACATCTGGATTATTAAGCCCGGGCCCAAGTCCCGGGCCGAGACAT	242
DB	653	GGAAGGGGATCGCAACATCTGSAATCGTGAAGCAGGAGCAAGTCCCGTGGACGAGCAT	712
QY	243	AGTGTGCATGGACCGGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTTC	302
DB	713	CATGTGATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTAT	772
QY	303	CAGGGAACAAGTGGTGGTCCAGAAATACATCGAGACCCGCTGCTCATCTGTGACAC	362
DB	773	GAAGGACGGCAAGTGGTGGTGCAGAAATATATTGAGCGGCCCTCTCTATCTTTGGCAC	832
QY	363	CAAGTTGACATCAGACAGTGGTTCCTGTCAGGACTGSAACCCCTGACCATCTGGTT	422
DB	833	CAAGTTTGAACCTCAGACAGTGGTTCCTGGTAACTGATGGAACCCACTTACCGTGGTT	892
QY	423	CTACAAGGAGAGTTACTTTGGGTTCTCACTCAGCGCTTCTCCTGGACAAGCTGGACAG	482
DB	893	CTACCGGACAGTATATCGCTTTTCCAGCGAGCCCTTCTCCTGAAGAACCTGGACAA	952
QY	483	CGCCATCCACCTGTCAACAACCCGCTCCAGAAATACCTGGAAGAAATGATGTGGGCCCGAG	542
DB	953	CTCAGTGCACTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCTCCATCGGCA	1012
QY	543	CCCCCTGCTGCCGACACACATGTGGACCAAGCTTCCAGGAGTTCAGGAGTACCTGGACGG	602
DB	1013	TCCACTGCTTCCGCCAGACAACTGTGGTCTAGCCAGAGGTTCCAGGCCCACTTGCAGGA	1072
QY	603	CCAGGCGCGTGGCGCGTGTGGGGCAGCGTCACTTACCCGCTCCATGAAGAGGCCATCGC	662
DB	1073	GATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGCTGGCATGAGGATGCTGTGAT	1132
QY	663	CCACGCCATGAAGTGGCGCCAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTA	722
DB	1133	CCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTCCGGAAGGCCAGCTTTGAGCTCTA	1192
QY	723	CGGGGCTGACTTCTGCTTCTGGGAGGAGCTTCAGGCGCTTGGCTGATCGAGATCAATTCAG	782
DB	1193	TGGCGCTGACTTCTGCTTCTGGGAGGAGCTTCCAGGCCCTGGCTGATTTGAGATCAACGCCAG	1252
QY	783	CCCCACCATGCACCCGCTCCAGCGCGTCCAGGCCAGCTGTGTGTCACAGGTGCAGGAGGA	842
DB	1253	CCCCAGATGGCACTTCCAGAGCTGCTGCTGCCGCTCTGTGCTGGCGTGCAGAGCTGA	1312
QY	843	CACCATCAAGGTGGCGCTGGACCGCA 868	
DB	1313	CACCTCGCGTGGTCAATTGACCGCA 1338	
	RESULT 12		
	AD573168		
	ID	AD573168 standard; cDNA; 2538 BP.	
XX	XX	AD573168;	
AC	AD573168;		
XX	XX	18-NOV-2004 (first entry)	
DT	DT	Human kidney tumour specific cDNA, SEQ ID 1765.	
DE	DE	Human; ss; kidney tumour; kidney cancer; cytostatic; gene therapy;	
XX	XX	Human; ss; kidney tumour; kidney cancer; cytostatic; gene therapy;	

KW T-cell; immune response.
XX Homo sapiens.
XX US2003109434-A1.
XX 12-JUN-2003.
XX 19-MAR-2002; 2002US-00102524.
XX 19-MAR-2001; 2001US-0277245P.
XX 21-DEC-2001; 2001US-0343340P.
XX (CORI-) CORIXA CORP.
XX Algate PA, Mannion J, Gaiger A, Gordon B, Harlocker SL;
XX WPI; 2002-759855/82.
XX New isolated polynucleotides and polypeptides, useful for detecting the
XX presence of, and treating cancer, particularly kidney cancer by
XX stimulating T-cells specific for a tumor protein, and stimulating immune
XX response in a patient.
XX Claim 1; SEQ ID NO 1765; 78pp; English.
XX The invention relates to a new isolated polynucleotide (a Human kidney
XX tumour specific cDNA) comprising any one of the 1855 sequences identified
XX in the specification (or their complements, degenerate variants,
XX sequences consisting of at least 20 contiguous residues them, sequences
XX that hybridise to them under highly stringent conditions or sequences
XX having at least 75 or 90% sequence identity to the 1855 sequences. Also
XX included are detecting/determining the presence of cancer in a patient,
XX stimulating an immune response in a patient; treating kidney cancer in a
XX patient; an isolated polypeptide encoded by one of the 1855 sequences, an
XX expression vector comprising the polynucleotide operably linked to an
XX expression control sequence, a host cell transformed/transfected with the
XX vector, an isolated antibody (or its antigen-binding fragment) that
XX specifically binds to the protein, a fusion protein comprising at least
XX one of the proteins, stimulating and/or expanding T-cells specific for a
XX tumour protein, an isolated T-cell population comprising the T-cells, a
XX composition comprising a first component (such as a carrier or
XX immunostimulant) and a second component (comprising one of the
XX polynucleotides, the polypeptides, an antibody, T-cell or an antigen-
XX presenting cell that expresses the polynucleotide) and a diagnostic kit
XX comprising at least one of the oligonucleotides, or at least one antibody
XX and a detection reagent comprising a reporter group. The polynucleotides,
XX polypeptides, antibodies and antigen-presenting cells are useful for
XX detecting the presence of, and treating cancer, particularly kidney
XX cancer by stimulating and/or expanding T-cells specific for a tumour
XX protein, and stimulating immune response in a patient. The present
XX sequence is one of the Human kidney tumour specific cDNAs. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html?docID=20030109434.
XX
SQ Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;
Query Match 44.0%; Score 398.8; DB 7; Length 2538;
Best Local Similarity 66.3%; Pred.No.1.3e-85;
Matches 5/4; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY 3 GGACATCGACAGTCTGAGGAGTCCGCTGAGGAGACCTCTAGGCGGAGTGGAGGACCT 62
Db 894 GGACATCGACAGGACCTGAGGCGCCGCTGTACCTCACCCTGAGGCGTGTCTCTTT 953
QY 63 GACCCAGAGTACTACTCTCTCTGTCGCGAGTCTTCTCTCAATTCGAAGAAATTA 122
Db 954 CCTCCAGCGCTACTACCAAGTGTCTCCAGAGGGGCGAAGAACTCAGGCACTCA 1013
QY 123 CTTTTCGAGTCCAGGCTCTGTGATAGATACGTCGTGAACCTTCAGACGACAT 182
Db 1014 GGTCCAGCGCTGTGAGGACATCTCTGAGCAGCTGCAGGCGGTGGTACCCCATAGACAT 1073

QY 183 TGACGGGCTCGGAAACATCTGGATTATAAGCCCGCGCCAAAGTCCCGGGCGGAGACAT 242
Db 1074 GGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAGTCCCGGGAGGAGCAT 1133
QY 243 AGTGTGCATGGACCGGTGTGGAGGAGATCTCTGAGCTGGCAGCTGCAGACCACTCTTTC 302
Db 1134 CATGTGCATGGACCACTCTGGAGGAGATGCTGAAGCTGTGTGAACGGCAACCCCGTGTGTAT 1193
QY 303 CAGGGACAAACAGTGGTGTCCAGAAAGTACATCGAGAGCGCGCTGCTCATCTGTGACAC 362
Db 1194 GAAGGACGCAAGTGGTGTGCAGAAAGTATATTGAGCGGCCCCCTCTCTCATCTTTGGCAC 1253
QY 363 CAAAGTTTCGACATCAGACAGTGTCTCTCTGTCACGGAAGTGGAAACCCCTGACCATCTGGTT 422
Db 1254 CAAAGTTTGACCTCAGACAGTGTCTCTCTGTTAACTGACTGGAACCCCACTTACCGTGTGT 1313
QY 423 CTACAGGAGAGTACTTACCTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGACAG 482
Db 1314 CTACCGCGACAGCTATATCCGCTTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTGGACAA 1373
QY 483 CGCCATCCACCTGTGCAACAAACGCGTCCAGAAAGTACCTGGAAGAAATGATGTGGCCGCGAG 542
Db 1374 CTCAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCCATCGGCA 1433
QY 543 CCCCTGCTGCCCCGACACATGTGGACCAAGCTGAGGAGTTCAGGAGTACCTGCGAGCG 602
Db 1434 TCCACTGCTTCCGCGCAGACAAACATGTGTGTCTAGCCAGAGGTTCCAGGCCCACTGCGAGGA 1493
QY 603 CAGGGCGGTGGCGCGGTGTGGGCGAGCTCATCTACCGTCCATGGAAGAGCCCATCGC 662
Db 1494 GATGGTGGCCCCAAATGTCTTGGTCCACCATCATCTGCGCTGCGCATGAAGGATGCTGTGAT 1553
QY 663 CCACGCCATGAAGTGGCGCCAGGACACCGTGGAGCTCGCAAGAACAGCTTTGAGCTCTA 722
Db 1554 CCACGCACTTACAGCTCCAGGACACCGTGCAGTGTGGAAGGCGCAGCTTTGAGCTCTA 1613
QY 723 CGGGGCTGACTTTCGCTCTTGGGAGGAGTCTCAGGCGCTGGCTGTGATCGAGATCAATTCAG 782
Db 1614 TGGCGCTGACTTTCGCTTTCGGGGAGGAGCTTCCAGCCCTGGCTGATTGAGATCAAGCCAG 1673
QY 783 CCCACCATGACCCGTCACCGCCGTCACGCGCCAGCTGTGTGACAGGTCAGGAGGA 842
Db 1674 CCCACGATGGCACCCTCCACAGCAGTCACTCCCGGCTCTGTGCTGGCGTGCAGACTGA 1733
QY 843 CACCATCAAGTGGCGCGTGACCGCA 868
Db 1734 CACCTGCGGTGTGTCATTGACCGGA 1759
RESULT 13
ACN37881
ID ACN37881 standard; cDNA; 2553 BP.
XX AC ACN37881;
XX
DT 18-NOV-2004 (first entry)
XX
XX Tumour-associated antigenic target (TAT) cDNA DNA324273, SEQ ID NO:1057.
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.
XX Homo sapiens.
XX WO2004030615-A2.
XX 15-APR-2004.

XX 29-SEP-2003; 2003WO-US028547.
PF
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
PI Wu TD, Zhang Z, Zhou Y;
XX
XX
DR WPI; 2004-347921/32.
XX
DR P-PSDB; ABM80420.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
XX Claim 1; SEQ ID NO 1057; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
XX
SQ Sequence 2553 BP; 575 A; 716 C; 733 G; 529 T; 0 U; 0 Other;

Query Match 44.08; Score 398.8; DB 13; Length 2553;
Best Local Similarity 66.3%; Pred. No. 1.3e-85;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 3 GGACATCGACAGCTGACAGATGCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGACCT 62
Db |||||
QY 63 GACCCAGCAGTACTACTCCCTCGTTCTATGCGGATGCTTTCATCTCCAAATTCAGAAATTA 122
Db |||||
QY 954 CCTCCAGCGCTACTACCAAGTGGTCCACGAAGGGGCGAGAACTCAGGCACCTCGACACTCA 1013
QY 123 CTTTTCAGTGCAGGCTCTGTGTATAGATACAGTCTGTGACCTCAGACGGACAT 182
Db |||||
QY 1014 GGTCACGCGTGTGAGGACATCTCTGAGCAGCTGCAGGCGCGTGTGTACCCAGATAGACAT 1073
QY 183 TGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGCAAGTCCCGGCGCGAGACAT 242
Db |||||
QY 1074 GGAAGGGATGCGAACATCTGGATCGTGAAGCCAGGACCAAGTCCCGCGGAGAGGCAT 1133
QY 243 AGTGTGATGAGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACCTCTTTC 302
Db |||||
QY 1134 CATGTGATGAGACCACTGGAGGAGATGCTGAAGCTGTGAACGGCAACCCCGTGTGTAT 1193
QY 303 CAGGGAACAAGTGGTGTGCGAAGTATCATGAGACCGCGCTGTCTATCTGTGACAC 362
Db |||||
QY 1194 GAAGGACGGCAAGTGGTGTGCAAGATATATTGAGCGGCGCCCTCTCATCTTTGGCAC 1253
QY 363 CAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGGAACCCCTGACCATCTGGTT 422
Db |||||
QY 1254 CAAGTTTGACCTCAGACAGTGGTTCCTCGTAACTGACTGGAACCCCACTTACCGTGTGGTT 1313

QY 423 CTACAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGACAAAGCTGGACAG 482
Db |||||
QY 1314 CTACCGCGACAGCTATATCCGCTTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTTGSACAA 1373
QY 483 CGCATCTCCACTGTGCAACAACGCGTCCAGAGAGTACTCTGAAGAATGATGTGGGCCCGCAG 542
Db |||||
QY 1374 CTAGTGCACCTGTGCAACAACCTCCATCCAGAGACCTCGAGAACTCATGCGCATCGCA 1433
QY 543 CCCCCTGCTGCCCGCACACAACATGTGGACCAAGTCCAGGATTCAGAGATCACTTCGACGC 602
Db |||||
QY 1434 TCCACTGCTTCCGCCAGACAACATGTGGTCTAGCCAGAGGTTCAGGCCCCACCTGCGAGGA 1493
QY 603 CAGAGGCGGTGGCGCGTGTGGGCGACGTCATCTACCCGTCATGAAGAGGCCATCGC 662
Db |||||
QY 1494 GATGGGTGCCCAATGCTTTGGTCCACCATCATCTGTCGCTGGCATGAAGGATGCTGTGAT 1553
QY 663 CCACGCCATGAAGGTGGCGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTCTA 722
Db |||||
QY 1554 CCAGGCACTTCAGACCTCCAGGACACCGTGCAGTGTCCGAAGGCCAGCTTTGAGCTCTA 1613
QY 723 CGGGGCTGACTTCTGCTTCTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTCAG 782
Db |||||
QY 1614 TGGCGTGAATCTGCTGTTGGGGGAGGACTTCCAGCCCTGCTGATTGAGATCAACGCCAG 1673
QY 783 CCCACCATGCAACCCGTCCAGCGGTGTCAGGCGCCAGCTGTGTGCACAGGTGCGAGGAGGA 842
Db |||||
QY 1674 CCCACGATGGCAACCCCTCCACAGCAGTCACTGCGCCGCTCTGTGCTGCGCTGCAAGCTGA 1733
QY 843 CACCATCAAGTGGCGCTGGACCGCA 868
Db |||||
QY 1734 CACCTGCGGTGTGTCATTGACCGGA 1759

RESULT 14
ADM03081
ID ADM03081 standard; cDNA; 2848 BP.
XX
AC ADM03081;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:1766.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI; 2003-723558/69.
DR
DR P-PSDB; ADM05524.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 1766; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene

This Page Blank (uspto)

Result No.	Score	Query			DB	ID	Description
		Match	Length	†			
C	1	817	90.2	817	6	CQ724907	Sequence
	2	427	47.1	101270	9	HS3555CJ18	AL022327 Human DNA
	3	400.4	44.2	4238	9	HSR805098	AL833939 Homo sapi
	4	398.8	44.0	2326	6	BD158727	BD158727 Primer fo
	5	398.8	44.0	2326	6	AX881015	Sequence
	6	398.8	44.0	2326	6	AQ023960	AX881015 Sequence
	7	398.8	44.0	2380	6	AX301197	AX023960 Homo sapi
	8	398.8	44.0	2553	9	HSR800637	AX301197 Sequence
	9	398.8	44.0	2848	6	AX834642	AL096725 Homo sapi
	10	398.8	44.0	2848	6	AX972326	AX834642 Sequence
C	11	398.8	44.0	3001	6	AR339003	AK097236 Homo sapi
	12	376.8	41.6	1684	9	AF078842	AR339003 Sequence
	13	366.4	40.4	1897	10	BC006830	AF078842 Homo sapi
	14	281.6	31.1	163783	10	AC119959	BC006830 Mus muscu
	15	281.6	31.1	201420	2	AC117700	AC119959 Mus muscu
	16	274.2	30.3	205949	2	AC349440	AC117700 Mus muscu
	17	274.2	30.3	218249	2	AC097425	AC349440 Rattus no
	18	212.2	23.4	5282	6	CQ842940	AC097425 Rattus no
	19	212.2	23.4	5282	9	AK125875	CQ842940 Sequence
	20	212.2	23.4	5282	9	AK125875	AK125875 Homo sapi

Db 181 CTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGAACAAGTGGTGTGCAGAA 240
QY 330 GTATATCGAGAGCCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCT 389
Db 241 GTATATCGAGAGCCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCT 300
QY 390 CGTACGCGACTGGAACCCCTGACCATCTGTTTCTACAAGGAGGTTACTTGCGGTTCTC 449
Db 301 CGTACGCGACTGGAACCCCTGACCATCTGTTTCTACAAGGAGGTTACTTGCGGTTCTC 360
QY 450 AACTAGCGCTTCTCCCTGGACAGCTGCAGACGCCATCCACCTGTGCAACAAGCCGT 509
Db 361 AACTAGCGCTTCTCCCTGGACAGCTGCAGACGCCATCCACCTGTGCAACAAGCCGT 420
QY 510 CAGAAGTAGTGAAGAAATGTTGGCGCGCAGCCCTGCTGCCGCGCACACAATGTG 569
Db 421 CAGAAGTAGTGAAGAAATGTTGGCGCGCAGCCCTGCTGCCGCGCACACAATGTG 480
QY 570 GACCAGCACCAGTTCCAGGAGTACCTGCAGCGCCAGGGCCCGTGGCGCGTGGGGCAG 629
Db 481 GACCAGCACCAGTTCCAGGAGTACCTGCAGCGCCAGGGCCCGTGGCGCGTGGGGCAG 540
QY 630 CGTCATCTACCGTGCATGAAGAGCCATCGCCGACAGGATGAAGTGGGCCAGGACCA 689
Db 541 CGTCATCTACCGTGCATGAAGAGCCATCGCCGACAGGATGAAGTGGGCCAGGACCA 600
QY 690 CTGGAGCTCGCAAGACAGCTTTCAGCTCTACGGGGCTGACTTCGCTTGGGAGGCA 749
Db 601 CTGGAGCTCGCAAGACAGCTTTCAGCTCTACGGGGCTGACTTCGCTTGGGAGGCA 660
QY 750 CTTGAGCCCTGGCTGATCGAGATCAATTCAGCCGCCACCATGCACCCGTCACGCGGT 809
Db 661 CTTGAGCCCTGGCTGATCGAGATCAATTCAGCCGCCACCATGCACCCGTCACGCGGT 720
QY 810 CAGGCGCAGCTGTGTGACAGGTGACAGGACACCATCAGTGGCGCGTGCACCGCAG 869
Db 721 CAGGCGCAGCTGTGTGACAGGTGACAGGACACCATCAGTGGCGCGTGCACCGCAG 780
QY 870 CTGTGACATCGCAACTTCGAGCTCTGTGGAGGCGAG 906
Db 781 CTGTGACATCGCAACTTCGAGCTCTGTGGAGGCGAG 817

RESULT 2
HS355C18/c
LOCUS
DEFINITION
Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3
Contains the KIAA0027 gene, ESTs, STSs, GSSs and seven putative CpG islands, complete sequence.
ACCESSION
AL022327
VERSION
AL022327.17 GI:5304851
KEYWORDS
HTG; CpG island; KIAA0027.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 101270)
Cobley V.
Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Jun 30, 1999 this sequence version replaced gi:5262834.
COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep -----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RP3-355C18 is from the library RPCI-3 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-355C18 The true
left end of clone RP5-89814 is at 5390 in this sequence.
FEATURES
Location/Qualifiers
1..101270
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="22"
/map="q13.3"
/clone="RP3-355C18"
/clone_lib="RPCI-3"
1..1038
/note="match: STS: Em:AL022483"
/complement(1..242)
/note="AluJo repeat: matches 1..237 of consensus"
358..484
/note="L1M4 repeat: matches 829..959 of consensus"
485..752
/note="AluSx repeat: matches 2..305 of consensus"
795..1095
/note="AluSx repeat: matches 1..299 of consensus"
1096..1165
/note="L1M4 repeat: matches 960..1031 of consensus"
1166..1563
/note="WSTB1 repeat: matches 4..432 of consensus"
1564..1750
/note="L1M4 repeat: matches 1031..1203 of consensus"
1757..2038
/note="AluSx repeat: matches 1..282 of consensus"
complement(2054..3342)
/note="SVA repeat: matches 1..1386 of consensus"
3343..3359
/note="2.8 copies 6 mer TCTCTT 34% conserved"
3360..3417
/note="L1TR39-int repeat: matches 3904..3961 of consensus"
3418..3577
/note="L1M4 repeat: matches 1203..1363 of consensus"
3578..3861
/note="AluJo repeat: matches 1..300 of consensus"
3862..4334
/note="L1M4 repeat: matches 1363..1830 of consensus"
4341..4489
/note="L1M4 repeat: matches 2065..2311 of consensus"
complement(4492..4840)
/note="MER41 repeat: matches 96..472 of consensus"
4990..4999
/note="2.5 copies 4 mer CCTG 20% conserved"
complement(5142..5300)
/note="AluJb repeat: matches 137..295 of consensus"

```
repeat_region complement(5315..5623)
repeat_region /note="AluX repeat: matches 3..312 of consensus"
repeat_region 5773..5863
repeat_region /note="MR81 repeat: matches 17..112 of consensus"
repeat_region 5917..5935
repeat_region /note="2.7 copies 7 mer GGGCCCA 29% conserved"
repeat_region 6130..6140
repeat_region /note="3.7 copies 3 mer GGT 22% conserved"
repeat_region 6139..6148
repeat_region /note="2.5 copies 4 mer GGAG 20% conserved"
repeat_region 6229..6240
repeat_region /note="2.0 copies 6 mer CAGCAC 24% conserved"
repeat_region 6573..6590
repeat_region /note="3.0 copies 6 mer CAGGAG 27% conserved"
repeat_region 6691..6700
repeat_region /note="2.0 copies 5 mer TGTGC 20% conserved"
repeat_region 6731..6862
repeat_region /note="3.3 copies 40 mer
repeat_region CCAGGTAGGGAGGACCTGGCGAGTGGCAGCTGCA 219% conserved"
repeat_region 7338..7356
repeat_region /note="2.7 copies 7 mer GCCCAC 29% conserved"
repeat_region 7636..7664
repeat_region /note="2.1 copies 14 mer CCCAGAGTGGCC 49% conserved"
repeat_region 7893..7906
repeat_region /note="2.0 copies 7 mer TGCACC 28% conserved"
repeat_region 7969..7980
repeat_region /note="2.0 copies 6 mer CCATGA 24% conserved"
repeat_region 8306..8316
repeat_region /note="2.8 copies 4 mer GTGG 22% conserved"
repeat_region 8399..8414
repeat_region /note="3.2 copies 5 mer GAGGG 23% conserved"
repeat_region 8840..8851
repeat_region /note="2.4 copies 5 mer AGCTG 24% conserved"
repeat_region 8939..8958
repeat_region /note="2.2 copies 9 mer CCCGCCCA 40% conserved"
repeat_region 9125..9135
repeat_region /note="2.2 copies 5 mer GTCT 22% conserved"
repeat_region 9246..9256
repeat_region /note="3.7 copies 3 mer GGA 22% conserved"
repeat_region 9263..9274
repeat_region /note="2.0 copies 6 mer CAGTGA 24% conserved"
repeat_region 9373..9383
repeat_region /note="2.2 copies 5 mer CCTGA 22% conserved"
repeat_region 9434..9444
repeat_region /note="2.2 copies 5 mer CAGC 22% conserved"
repeat_region 9616..9631
repeat_region /note="2.3 copies 7 mer GGGATG 32% conserved"
repeat_region 10067..10076
repeat_region /note="2.5 copies 4 mer GGCA 20% conserved"
repeat_region 10245..10254
repeat_region /note="2.5 copies 4 mer CTT 20% conserved"
repeat_region 10496..10531
repeat_region /note="2.0 copies 18 mer GCGAGGACAGGACAGG 63% conserved"
repeat_region 10498..10531
repeat_region /note="2.8 copies 12 mer CAGGGCAGGA 59% conserved"
repeat_region 10745..10756
repeat_region /note="2.0 copies 6 mer GCTGG 24% conserved"
repeat_region 10889..10903
repeat_region /note="2.1 copies 7 mer CCGGCC 30% conserved"
repeat_region 11166..11193
repeat_region /note="3.5 copies 8 mer TGTGTGA 33% conserved"
repeat_region 11177..11211
repeat_region /note="5.8 copies 6 mer GTGCAT 38% conserved"
repeat_region 11250..11261
repeat_region /note="2.0 copies 6 mer ATGGGT 24% conserved"
repeat_region 11273..11282
repeat_region /note="2.5 copies 4 mer TGCC 20% conserved"
repeat_region 11293..11309
repeat_region /note="2.8 copies 6 mer TGTGA 25% conserved"
repeat_region 11311..11411
repeat_region /note="2.5 copies 40 mer
repeat_region TGCATGGGTGATGGTATGGTGTGGTGTGGTGTGCCT 202% conserved"
11323..11334
```

```
repeat_region /note="2.0 copies 6 mer ATGGGT 24% conserved"
11346..11355
repeat_region /note="2.5 copies 4 mer TGCC 20% conserved"
11368..11379
repeat_region /note="2.0 copies 6 mer ATGGGT 24% conserved"
11391..11400
repeat_region /note="2.5 copies 4 mer TGCC 20% conserved"
11431..11683
repeat_region /note="7.0 copies 36 mer
repeat_region TGCATGTATGGGTGTGTATGTGATGCGTGGTGGTG 108% conserved"
11451..11490
repeat_region /note="2.9 copies 14 mer TGTGATGGTGGC 71% conserved"
11452..11477
repeat_region /note="3.2 copies 8 mer GTGCATGG 29% conserved"
11485..11498
repeat_region /note="2.3 copies 6 mer TGGGTG 28% conserved"
11513..11523
repeat_region /note="2.8 copies 4 mer TGCC 22% conserved"
11521
misc_feature
11521
repeat_region /note="Tandem repeat. Assembly not confirmed by digest"
11531..11562
repeat_region /note="5.3 copies 6 mer TGTGCA 32% conserved"
11536..11635
repeat_region /note="6.2 copies 16 mer GTGTGATGTGTATGG 82% conserved"
11595..11604
repeat_region /note="5.0 copies 2 mer TG 20% conserved"
11679..11690
repeat_region

Query Match 47.1% Score 427; DB 9; Length 101270;
Best Local Similarity 100.0%; Pred. No. 1.7e-68;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 CAGCGCCATCCACCTGTGCAACAAACCCGCTCCAGAAGTACCTGAAAGATGATGGCCG 539
DB 44821 CAGCGCCATCCACCTGTGCAACAAACCCGCTCCAGAAGTACCTGAAAGATGATGGCCG 44762

QY 540 CAGCCCCCTGCTGCCCGCACAAACATGTGGACAGCACAGCTTCCAGAGTACCTGCA 599
DB 44761 CAGCCCCCTGCTGCCCGCACAAACATGTGGACAGCACAGCTTCCAGAGTACCTGCA 44702

QY 600 GCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTATCTACCCGCTCCATGAAGGCCAT 659
DB 44701 GCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTATCTACCCGCTCCATGAAGGCCAT 44642

QY 660 CCGCCACGCCATGAAGGTGGCCCGCAGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 719
DB 44641 CCGCCACGCCATGAAGGTGGCCCGCAGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 44582

QY 720 CTACGGGGTGACTTCTGCTCTTGGGAGGACCTTCAGGCGCTGGCTGATCGAGATCAATTC 779
DB 44581 CTACGGGGTGACTTCTGCTCTTGGGAGGACCTTCAGGCGCTGGCTGATCGAGATCAATTC 44522

QY 780 CAGCCCCCAGCAGCAGCCCGCTCCACGCGGTCCAGGCGCCAGCTGTGTGCACAGGTGCAGGA 839
DB 44521 CAGCCCCCAGCAGCAGCCCGCTCCACGCGGTCCAGGCGCCAGCTGTGTGCACAGGTGCAGGA 44462

QY 840 GGACACCATCAAGGTGGCCGCTGGAACCGCAGCTGTGTGATCGGCAACTTCGAGCTCTGTG 899
DB 44461 GGACACCATCAAGGTGGCCGCTGGAACCGCAGCTGTGTGATCGGCAACTTCGAGCTCTGTG 44402

QY 900 GAGGCAG 906
DB 44401 GAGGCAG 44395

RESULT 3
LOCUS HSM805098 4238 bp mRNA linear PRI 12-JUL-2002
DEFINITION Homo sapiens mRNA; cDNA DKFZp586B0320 (from clone DKFZp586B0320).
ACCESSION AL833939
VERSION AL833939.1 GI:21739506
KEYWORDS
SOURCE Homo sapiens (human)
```



```

FEATURES             FT  CDS           Location/Qualifiers
     source            1..2326
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
ORIGIN
Query Match          44.0%; Score 398.8; DB 6; Length 2326;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY   3  GGACATCGACAGCTCAGCAGATCCGCTGGAGGACCTCCTCAGCGCCGAGTGGAGGACCT 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   732 GGACATCGACAGGACCTGAGGCGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTCTT 791
QY   63  GACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAGAAATTA 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   792 CCTCAGCGCTACTACCAAGTGTCCACGAAGGGCAGAACTCAGGCACCTCGACACTCA 851
QY   123 CTTTTCGAGTGCAGGCTCTCTCTGAATAGATCAGCTCTGTGAACCTCAGACGACAT 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   852 GGTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCTGTACCCAGATAGACAT 911
QY   183 TCACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGGCAAGTCCCGGGCCGAGACAT 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   912 GGAAGGGGATCCCAACATCTGGATCGTGAAGCCAGAGCCAAGTCCCGGGAGCGGACAT 971
QY   243 AGTGTGCATGGACCGTGTGGAGGAGATCTGGAGCTGGCAGCTGCAGACACCCCTCTTTC 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   972 CATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGTGAT 1031
QY   303 CAGGAGCAACAAGTGGGTGGTCCAGAAATCATCGAGACCGCGCTGCTCATCTGTGACAC 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1032 GAAGGACGGCAAGTGGGTGGTGCAGAAATATATTGAGCGGCCCTCTCTATCTTTGGCAC 1091
QY   363 CAAGTTTCGATCAGACAGTGGTTCTCTCGTCAGGACTGGAACCCCTGACCATCTGTT 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1092 CAAGTTTGACCTCAGACAGTGGTTCTCTGTTAACTGACTGGAACCCACTTACCGGTGGTT 1151
QY   423 CTACAAGGAGAGTTACTTTCGGTCTTCAACTCAGCGCTTCTCCCTCGACAAAGCTGACAG 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1152 CTACCGGACAGTATATCGCTTTTCCAGCGAGCCCTTCTCCTGAAGAACCTGGACAA 1211
QY   483 CGCCATCACTGTGCAACAAACCCCTCCAGAAATACCTGAAGAAATGATGTGGGCGGACAG 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1212 CTCAAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACCTGGAGAACTCATGCCATCGGCA 1271
QY   543 CCCCCTGCTGCCGACACAAACATGTGGACCAACAGTTCAGGAGTTCAGGAGTACTCTGACGG 602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1272 TCACCTGCTTCCGCCAGACAAACATGTGGTCTAGCCAGAGGTTCCAGGCGCCACCTGCAGGA 1331
QY   603 CCAGGCGCGTGGCGCGTGTGGGGGACGCTCATCTACCCGCTCCATGAAGAAGCCATCGC 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1332 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGTCCTGGCATGAAGATGCTGTGAT 1391
QY   663 CCACGCGCATGAAGGTGGGCCAGGACCAACGCTGGAGCCCTCGCAAGAAACAGCTTTGAGCTCTA 722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1392 CCACGCACTTCAGACCTCCAGGACACCCGTGAGTGTGCGAAGGCCAGCTTTGAGCTCTA 1451
QY   723 CGGGGCTGACTTTCGTTCTTGGGAGGACCTTCAGGCCCTGGCTGATCGAGATCAATCCAG 782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1452 TGGCGTGTACTTTCGTTCTGGGGAGGACTTTCAGGCCCTGGCTGATTGAGATCAACGCCAG 1511
QY   783 CCCACCATGCACCCGCTCAACGCGGTTCAGCGGCCAGCTGTGTGACAGCTGCAGGAGGA 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1512 CCCACGATGGCAACCTTCACAGCACTCACTGCCCCGGCTCTGTCTGGGTGCGAGCTGA 1571
QY   843 CACCATCAAGGTGGCGGTGGACCGCA 868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1572 CACCCTGCGGTGGTCAATTGACCGGA 1597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 5
AX881015
LOCUS             AX881015             2326 bp             DNA             linear             PAT 17-DEC-2003
DEFINITION       Sequence 15920 from Patent EP1074617.
ACCESSION       AX881015
VERSION         AX881015.1             GI:40035751
KEYWORDS
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1. Ota, T., Isegai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
    Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, I.
    Primers for synthesizing full-length cDNA and their use
    Patent: EP 1074617-A 15920 07-FEB-2001;
    Research Association for Biotechnology (JP)
FEATURES
     source            1..2326
                        /organism="Homo sapiens"
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:9606"
                        94..>2325
                        /note="unnamed protein product"
                        /codon_start=1
                        /protein_id="CAE92353.1"
                        /db_xref="GI:40035752"
                        /translation="NNLRNNAKIYVERAVKQKKIFTIOGYPVIRCLRRRWGVEKMM
VHRSGPTLLPQKDLDDSSAMGSDTDEDEDEDEFQPSQLPDPDLKLFDDLDGTH
ALMVGICLNRLNLPWFDEVANSFPRCYCLGAEDDKAFIDFMTLTAARNVLKLVVK
SWKSYPTQAVEEESAGKQPKQKPNVLSVPEFVDEALCAEEYLSNLAHMDIDKD
LSAPILTPGWSLFLQYVYVHGAELRLDHTQVRCEDILQQLOAVPQIDMGGD
RNINWFKPAKSGRIGIMCMHLEMLKLVNGNPVVMKDGWVQVYKIERPLIFGTH
FLDLQFLVTDWNPGLTVMFDRSYIRFTQPSFLKNDNSVHLKNISIOQVOCRKASR
HLLPDDNNWSSQFOAHLQENAPNAWSTIIVPGMKDAVIALQTSQDTVOCRKASR
ELYGADFVGEDFQWLLIENASPTMAPSTAVTARLCAGVQADTLRVVLDRLDNCD
TCAPHLIYQPAVEVPQYVIGIRLLVEGFTIKKPMAMCHRRMRGVRPAVPLLTQSGEA
RIHPSFLHTKAQPSHVLRHQGVLRHSLKLVGTAKLSTTGKALURTUPTKATVLSL
PNPLDFKVAPSLKPRKALCLRGQPLEVPCCLGKPSKSEQFLAPVGRSPKANSRP
DCDKPRACPMKRLSLPLPLVGTGTRRRGLGDMKLGKPLRLFPPTALVLDTPNKK
KQVK"
ORIGIN
Query Match          44.0%; Score 398.8; DB 6; Length 2326;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY   3  GGACATCGACAGCTCAGCAGATCCGCTGGAGGACCTCCTCAGCGCCGAGTGGAGGACCT 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   732 GGACATCGACAGGACCTGAGGCGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTCTT 791
QY   63  GACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAGAAATTA 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   792 CCTCAGCGCTACTACCAAGTGTCCACGAAGGGCAGAACTCAGGCACCTCGACACTCA 851
QY   123 CTTTTCGAGTGCAGGCTCTCTCTGAATAGATCAGCTCTGTGAACCTCAGACGACAT 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   852 GGTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCTGTACCCAGATAGACAT 911
QY   183 TCACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGGCAAGTCCCGGGCCGAGACAT 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   912 GGAAGGGGATCCCAACATCTGGATCGTGAAGCCAGAGCCAAGTCCCGGGAGCGGACAT 971
QY   243 AGTGTGCATGGACCGTGTGGAGGAGATCTGGAGCTGGCAGCTGCAGACACCCCTCTTTC 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   972 CATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGTGAT 1031
QY   303 CAGGAGCAACAAGTGGGTGGTCCAGAAATCATCGAGACCGCGCTGCTCATCTGTGACAC 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1032 GAAGGACGGCAAGTGGGTGGTGCAGAAATATATTGAGCGGCCCTCTCTATCTTTGGCAC 1091
QY   363 CAAGTTTCGATCAGACAGTGGTTCTCTCGTCAGGACTGGAACCCCTGACCATCTGTT 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 1092 CAAGTTTGACCTCAGACAGTGGTTCTCGTGAACCTGACGCCAACCCCACTTACCGTGTGGTT 1151
QY 423 CTACAAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTCGAACAAGCTGACAG 482
Db 1152 CTACCGGACAGCTATATCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACTGGACAA 1211
QY 483 CGGCATCCACTGTGTGAACAAACGCCCGTTCAGAAAGTACCTGAAGAATGATGTGGCGCGCAG 542
Db 1212 CTCAGTGCACCTGTGCAACAACACTCCATCCAGAAAGCACCTGGAGAACTCATGCCATCGCA 1271
QY 543 CCCCTGCTGCCCGCACACAAACATGTCGACCCAGCCAGCTTCCAGAGTACTCTGCGCG 602
Db 1272 TCACACTGCTTCCGCCAGAACATATGGTGTAGCCAGAGGTTCAGGCCCACTTCGAGGA 1331
QY 603 CCAGGCGCGTGGCGCGCTGTGGGCGACGCTCATCTACCGTTCATCAAGAAGCCATCGC 662
Db 1332 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGCTGCGCATGAAGATGCTGTGAT 1391
QY 663 CCACGCCATGAAGGTGGCCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTA 722
Db 1392 CCACGCACCTTCAGACCTCCAGGACACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTA 1451
QY 723 CGGGGCTGACTTCTCTTTGGGAGGACATTCAGGCCCTGCTGATCGAGATCAATTCAG 782
Db 1452 TGGCGTGTACTTCTGTTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAAGCCCA 1511
QY 783 CCCACCATGACCCCTGCACGCGGCTCAGCGGCCAGCTGTGTGACAGGTGCAGGAGGA 842
Db 1512 CCCACGATGGCACCTTCACAGCAGTCACTGCCCGGCTCTGTGCTGGGTGCAGACTGA 1571
QY 843 CACCATCAAGGTGGCGGTGGACCGCA 868
Db 1572 CACCCTGCGGTGGTCAATGACCGGA 1597

RESULT 6

AK023960
LOCUS Homo sapiens cDNA FLJ13898 fis, clone THYR01001738, weakly similar
DEFINITION to TUBULIN--TYROSINE LIGASE (EC 6.3.2.25).
ACCESSION AK023960
VERSION AK023960.1 GI:10436083
KEYWORDS oligo cappings; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yoshida, M., Hotta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togliya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hiroo, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigetate, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,

Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
2
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,
Tanase, T., Nomura, Y., Iogiya, S., Komai, F., Hara, R., Takeuchi, K.,
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2326)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry for Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
LOCATION/Qualifiers
1..2326
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYR01001738"
/tissue_type="thyroid gland"
/clone_lib="THYR01"
/note="cloning vector: pME18SFL3"
94..>2326
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14741.1"
/db_xref="GI:10436084"
/translation="MNRLNAKIYVERAVKQKIFTIQCVPIVIRCLRRRGWVRKRM
VHRSGTLLPQKLDSSAMGSDTDEDEDEEFQPSQDFDLDLLKFDLDLGTTH
ALMVGCLNRLNLPWFDEVDAMSPPRCVLCAGDEDDKAFIDFWLTAARNVLKLVK
SEWSPYICQAVEEASGDQPKQKPNVLSPFVDRALCALCEVLSNLAHMDIKD
LEAPYLYTQAVEEASGDQPKQKPNVLSPFVDRALCALCEVLSNLAHMDIKD
RNIMIVKPGASRGIMCMDEHLEMLKLVNPNVMDKRWVQKTIERPLLIPTGK
FDLRQWFLVTWNPLTVMFYRDSYTRFSTQPSFLNLDNSVHLNNSIQHLENSCHR
HPLLPDNNWSSQRFQHLQMGAPNASTIIPGMDAVIHALQSDQTVQCRKASF
ELYGADVFGEEDFQPLIENASPTMAPSTAVRLCAGVQADTLRVVDIRMLDRNC
TGAPFLIYKQPAVEVQVIGIRLLVEGFTIKKPMAMCHRMGVRPAVPLITORGSEA
RHHFSLTKAQLPSPHVRHQGVLRHQSKLVTKALSTGKALRILPTAKVFIISL
PPNDFKVPASITLKPRAKPAULCLRGPOLEVPCCCLPKSEQFLAPVGRSRPANSRP
DCDKPEACRACPMKRLSPLKPLFLVGTQRRRGLGMDKLGKLLRPTALVLDPTPNKK
QVKV"

FEATURES
source

1..2326
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYR01001738"
/tissue_type="thyroid gland"
/clone_lib="THYR01"
/note="cloning vector: pME18SFL3"
94..>2326
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14741.1"
/db_xref="GI:10436084"

CDS

1..2326
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14741.1"
/db_xref="GI:10436084"

ORIGIN

Query Match 44.0%; Score 398.8; DB 9; Length 2326;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY 3 GGACATCGACACGTCAGCAGATGCGCTGGAGGACCTCACTGAGGCGCGAGTGAGGACCT 62
Db 732 GGACATCGACAGGACCTTGAGGCGCCCGCTGTACTCACCCTGAGGCGTGTCTCTCT 791
QY 63 GACCCAGCAGTACTACTCCCTCGTTTCATGGGATCGTTTCATCTCAATTCAGAAATTA 122
Db 792 CCTCCAGCGCTACTACCAAGTGGTCCACGAGGGGAGAACTCAGGACCTCGACACTCA 851
QY 123 CTTTTCGAGTCCGAGGCTCTCTCTGAATAGATCATCGTCTGTGAACCTTCAGACGACAT 182

Db 852 GGTCCAGCGCTGTGAGGACATCTCTGAGCAGAGCTGTCAGGCGGTGGTATCCCCAGATAGACAT 911
Qy 183 TGACGGGCTCGGAACTATCTGGATTATTAAGCCCGCGCCAGTCCCGGGCCGAGACAT 242
Db 912 GGAAGGGATCGCAACTCTGGATCTGTAAGCCAGGAGCAAGTCCCGCGAGCAGGCAT 971
Qy 243 AGTGTGATGACCGCTGTGGAGGAGATCTTGGAGCTGCGAGCTGCGACACACCTCTTTTC 302
Db 972 CATGTGATGACCACTTGGAGGAGATCTGTAAGCTGGTGAACCGCAACCCCGTGGTGTAT 1031
Qy 303 CAGGGAACAAGTGGGTGGTCCAGAAAGTATCATCGAGACGCGCTGCTCATCTGTGTGACAC 362
Db 1032 GAAGGACGCAAGTGGGTGGTGCAGAAAGTATATTGAGCGGCCCTCTCATCTTTGGCAC 1091
Qy 363 CAAGTTGACATCAGACAGTGGTTCCTGTCGACGGAATGGAACCCCTGACCACTCTGGTT 422
Db 1092 CAAGTTTGAACCTCAGACAGTGGTTCCTGTTAACTGACTGGAACCACTTACCGTGTGGTT 1151
Qy 423 CTACAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAG 482
Db 1152 CTACCGGACAGCTATATCCGCTTTTCCAGCAGCCCTTCTCCCTGGAAGAACCTGGACAA 1211
Qy 483 CGCCATCCACTGTGCAACACGCGCTCCAGAAGTACTCTGAAGATGATGTGGCGCGCAG 542
Db 1212 CTAGTGCACCTGTGCAACAACTCCATCCAGAGCACTCGAGAACTCATGCCATCGCA 1271
Qy 543 CCCCTGCTGCGCGCAACAACATGTGTGAACAGCAGTTCAGAGTACCTGCGACG 602
Db 1272 TCCACTGCTTCCGCGACAGCAACATGTGTGTAGCAGAGGTTCCAGGCCCACTGCGAGGA 1331
Qy 603 CAGGGCGGTGGCGCGTGTGGGCGAGGTCATCTACCGTCCATGATGAAGAGCCATCGC 662
Db 1332 GATGGGTGCCCCAAATCTTGGTCCACATCATCGTGCCTGGCATGAAGATGCTGTGAT 1391
Qy 663 CCAGCCATGAAGTGGCGCCAGGACCACTGAGGAGCTTCGAAAGAAAGCTTTGAGCTCTA 722
Db 1392 CCAGCACTTCAGACCTCCAGGACACGCTGAGTGTGAGGAGCTTTGAGCTCTA 1451
Qy 723 CGGGCTGACTTCTGCTTGGGAGGAGCTTCAGGCGCTTGGCTGATGAGATCAATTCAG 782
Db 1452 TGGCGCTGACTTCTGTTTGGGAGGAGCTTCCAGCCCTGGCTGATTGAGATCAACGCCAG 1511
Qy 783 CCCACCATGACCCGTCAGCGCGTCCAGGCGCTGTCAGGCGCTGTCAGAGTGCAGGAGGA 842
Db 1512 CCCACCATGAGCCCTCCAGCAGAGTCACTGCCCCGCTCTGTGCTGGCGTGCAGAGCTGA 1571
Qy 843 CACCATCAAGTGGCGCTGGACCGCA 868
Db 1572 CACCTGCGGTGCTATTGACCGGA 1597

RESULT 7
AX301197 2380 bp DNA linear PAT 30-NOV-2001
LOCUS
DEFINITION Sequence 39 from Patent W00185942.
ACCESSION AX301197
VERSION AX301197.1 GI:17382288
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Yue,H., Tang,Y.T., Au-Young,J., Lu,D.A., Baughn,M.R., Hillman,J.L.,
Azimzai,Y., Lal,P., Yao,M.G., Bandman,O., Burford,N., Batra,S.,
Kearney,L. and Policky,J.L.
Cytoskeleton-associated proteins
Patent: WO 0185942-A 39 15-NOV-2001;
Incyte Genomics, Inc. (US)
Location/Qualifiers
1. .2380
/organism="Homo sapiens"
/mol_type="unassigned DNA"

/db xref="taxon:9606"
/note="Incyte ID No: 2156553CB1"

ORIGIN

Query Match 44.0%; Score 398.8; DB 6; Length 2380;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

Qy 3 GGACATCGACATCTGAGCAGATCGGTGGAGGACCTCCTCAGTGGCCGAGTGGGAGGACCT 62
Db 473 GGACATCGACAGGACCTGAGGCCCGCGTGTACCTCACCCCGAGGCGTGGTCCCTCTT 532
Qy 63 GACCGAGCATCTACTCTCCCTCGTTCATGSCGATGCTTTCATCTCCAATTCAAGAAATTA 122
Db 533 CTTTCCAGCGCTACTTACCAAGTGTCCAGAAAGGGGAGAACTCAGGCACCTCGACACTCA 592
Qy 123 CTTTTCGAGTGGCAGGCTCTGCTGAATCAAGTCTGTGTGAACCTCTCAGACGAGACAT 182
Db 593 GGTCCAGCGCTGTGAGGACATCTCGCAGCAGCTGCAGGCCGTGGTACCCAGATAGACAT 652
Qy 183 TGACGGGCTCCGAAACATCTGTGATTTATAAGCCCGCGGCCAAGTCCCGGGCCGAGACAT 242
Db 653 GGAAGGGGATCGCAACATCTGGATCTGTGAAGCAGGAGCAAGTCCCGTGGACGAGGCAT 712
Qy 243 AGTGTGCATGGACCGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCACTCTTTC 302
Db 713 CATGTGATGGAACCACTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGTAT 772
Qy 303 CAGGGAACAAGTGGGTGTCCAGAAAGTACATCGAGACCGCGCTGCTCATCTGTGACAC 362
Db 773 GAAGGACGGCAAGTGGGTGTGAGAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCAC 832
Qy 363 CAAGTTGACATCAGACAGTGGTTCCTGTCACGAGCTGGAAACCCCTGACCATCTGGTT 422
Db 833 CAAGTTTGAACCTCAGACAGTGGTTCCTGTGTAAGTGAAGTGAAGTGAAGTGAAGTGA 892
Qy 423 CTACAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCTCTGAGCAAGCTGGACAG 482
Db 893 CTACCGGACAGCTATATCCGCTTTTCCAGCAGCCCTCTCTCTGAAAGAACTTGAACAA 952
Qy 483 CGCCATCCACTGTGCAACAAACCGCTCCAGAAAGTACCTGAAGAAATGATGTGGCGCGCAG 542
Db 953 CTCAAGTGCACCTGTGCAACAACTCCATCCAGAGCAGCTGGAGAACTCATGCCATCGCA 1012
Qy 543 CCCCTGTGCTCCGCAACAACATGTGGAACAGACACAGGTTCCAGAGTACCTGCGACG 602
Db 1013 TCCACTGCTTCCGCGCAGACAAACATGTGGTCTAGCAGAGGTTCCAGGCCCACTGCGAGGA 1072
Qy 603 CAGGGCGGTGGCGCGTGTGGGCGAGCGTCACTACCGTCCATGAAGAGGCCATCGC 662
Db 1073 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGTCCTGGCATGAAGATGCTGTGAT 1132
Qy 663 CCACGCGCATGAAGGTGGCCCGCAGGACCAACGTTGGAGCCTCGCAAGAAACAGCTTTGAGCTCTA 722
Db 1133 CCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTA 1192
Qy 723 CGGGGCTGACTTCTGCTTGGGAGGAGTTCAGGCCCTGGCTGATCGAGATCAATTCAG 782
Db 1193 TGGCGCTGACTTCTGTTTGGGAGGACTTCCAGGCCCTGGCTGATTCAGATCAACGCCAG 1252
Qy 783 CCCACCATGACCCCTCCAGCGCGTCCAGGCCGCTCAGGCCAGCTGTGTGCAGAGTGCAGGAGGA 842
Db 1253 CCCACCATGAGCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGCAGAGCTGA 1312
Qy 843 CACCATCAAGTGGCGGTGGACCGCA 868
Db 1313 CACCTGCGGTGCTCATTGACCGGA 1338

RESULT 8
HSM800637 2553 bp mRNA linear PRI 18-FEB-2000
LOCUS
DEFINITION Homo sapiens mRNA; cDNA DKFZp434B103 (from clone DKFZp434B103);

partial cds.
AL096725
VERSION
AL096725.1 GI:5419858
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2553)
Wambutt R., Heubner D., Mewes H.W., Gassenhuber J. and Wiemann S.
Direct Submission
Submitted (16-JUN-1999) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp434B103) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
FEATURES
Location/Qualifiers
1..2553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp434B103"
/db_xref="taxon:9606"
/clone="DKFZp434B103"
/tissue_type="testis"
/dev_stage="adult"
DH10B; sites NotI + SalI
/dev_stage="adult"
1..2553
/gene="DKFZp434B103"
892..1950
/gene="DKFZp434B103"
/note="similarity to tubulin-tyrosine ligase"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB46375.1"
/db_xref="GI:5419859"
/db_xref="GOA:Q9Y4R7"
/db_xref="UniProt/Swiss-Prot:Q9Y4R7"
/translation="MDIDKLEAPLYLTPEKSLFLQRYQVYVHGAELRHLDTOVOR
CEDILQLOAVPOIDMEGRNIWIKPKAKSRGRICMCDHLEMLKLVNGNPVVMK
DKWVQKYLEPLIEGTFDLRQWFLVTDNPLTWYRDSYIRFSPQPSLNLD
NSVHLNNISIQHLENSCHRRFLPDDNWSSRFQHLQENGAPNWSIIIVPGKO
AVIHALQSDTVQCKRSFELYGADVFEGDFQPLWLEINASPTMAPSTAVTLCA
GYQADTLRVVIDRMLDRNCDTGAFELIYKQPVTTSPASTPRPSCLLIPMYSIDTRRSSD
DSTASWMLRPCRPOARP"
2511..2516
/gene="DKFZp434B103"
2528
/gene="DKFZp434B103"
polyA_signal
2511..2516
/gene="DKFZp434B103"
polyA_site
2528
/gene="DKFZp434B103"
ORIGIN
Query Match 44.0%; Score 398.8; DB 9; Length 2553;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY 3 GGACATCGACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGACCT 62
Db |||||
894 GGACATCGACGACGACCTGGAGGCCCGGTGTACTCCACCCGAGGCGTGGTCCCTCT 953
QY 63 GACCCAGCAGTACTACTCCTGTTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTA 122
Db |||||
954 CCTCCAGCGCTACTCAAGATGGTTCACGAAGGGGAGAACTCAGGCACCTCGACACTCA 1013
QY 123 CTTTTCGCGAGTGCAGGCTCTGCTGAATAGATCACTGTAACCCCTCAGACGAGCAT 182
Db |||||
1014 GGTCCAGCGCTGTGAGGACATCTCTGACGAGCTGCAGGCGGTGTACCCACGATAGACAT 1073

QY 183 TCACGGGCTCCGAAACATCTGGATTATAAAGCCCGCGCCAAAGTCCCGGGCGGAGACAT 242
Db |||||
1074 GGAAGGGGATCGCAACATCTGGATCGTGAAGCAGGAGCAAGTCCCGGCGAGGACAT 1133
QY 243 ACTGTGCATGGACCGTGTGGAGGAGATCTGGAGCTGGGAGCTGCAGACCACTCTTTC 302
Db |||||
1134 CATGTGCATGGACCACTCTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGTGTAT 1193
QY 303 CAGGAGCAACAAGTGGTGGTCCAGAAGTACATCGAGACCGCGCTGCTCATCTGTGACAC 362
Db |||||
1194 GAAGAGCGCAAGTGGTGGTGCAGAGATATATTGAGCGGCGCCCTCTCATCTTTGGCAC 1253
QY 363 CAAATTTCGACATCAGACAGTGTCTCTCTCAGGACTGGAACCCCTCACCATCTGGTT 422
Db |||||
1254 CAAATTTCGACATCAGACAGTGTCTCTCTCAGGACTGGAACCCCTCACCATCTGGTT 1313
QY 423 CTACAGGAGAGTACTTGGCGGTTCTCAACTCAGCGCTTCTCCCTGGAGCAAGCTGGACAG 482
Db |||||
1314 CTACCGCGACAGTATATCCGCTTTTCCACGAGCGCCCTCTCCCTGAAGAACCTGGACAA 1373
QY 483 CGCCATCCACCTGTGCAACAACCGCGTCCAGAACTACCTGGAAGAAATGATGTGGCGCGCAG 542
Db |||||
1374 CTCAGTGCACCTGTGCAACAACCTCCATCCAGAAGCACCTGGAGAACTCATGCCATCGCA 1433
QY 543 CCCCCTGTGTCGCGCACACAACATGTGGACCAAGCACAGGTTCACAGGATCTCTGAGCGG 602
Db |||||
1434 TCACCTGTCTCCGCGCAGACAACATGTGTCTAGCCAGAGGTTCCAGGCGCACCTGCAGGA 1493
QY 603 CCAGGCGCGTGGCGCGTGTGGGCGAGCGTCACTACCGTCCATGAGCAAGAGCCATCGC 662
Db |||||
1494 GATGGTGTGCCCAATGTGTGTTCACCATCATCTGTCCTGGCATGAAGATGCTGTGTAT 1553
QY 663 CCACGCCATGAAGTGGCGCGCAGGACCACTGGAGCGCTCCCAAGAAACAGCTTTGAGCTCTA 722
Db |||||
1554 CCACGCCATTCAGACCTCCAGGACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTA 1613
QY 723 CGGGGCTGACTTGTCTTGTGGGAGGAGCTTCAGGCGCTGCTGATGAGATCAATCCAG 782
Db |||||
1614 TGGCGCTGACTTGTCTTGTGGGAGGAGCTTCCAGCGCTTGGCTGATGAGATCAACGCCAG 1673
QY 783 CCGCACCATGACCGCTCCACCGCTCAGCGCCAGCTGCTGTGCACAGTGCAGAGTGCAGGAG 842
Db |||||
1674 CCGCACCATGACCGCTCCACCGCTCAGCGCCAGCTGCTGTGCACAGTGCAGAGTGCAGGAG 1733
QY 843 CACCATCAAGTGGCGCGTGGACCGCA 868
Db |||||
1734 CACCTCGCGTGGTTCATTGACCGGA 1759
RESULT 9
AX834642
LOCUS
Sequence 1766 from Patent EP1347046.
AX834642
ACCESSION
AX834642.1 GI:39920777
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Isogai T., Sugiyama T., Otsuki T., Wakamatsu A., Sato H., Ishii S.,
Yamamoto J.I., Isono Y., Hio Y., Otsuka K., Nagai K., Irie R.,
Tamechika I., Seki N., Yoshikawa T., Otsuka M., Nagahari K. and
Masuho Y.
Full-length cDNA sequences
Patent: EP 1347046-A 1766 24-SEP-2003;
Research Association for Biotechnology (JP)
TITLE
JOURNAL
FEATURES
Location/Qualifiers
1..2848
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN		Query Match		44.0%; Score 398.8; DB 6; Length 2848;		Best Local Similarity 66.0%; Pred. No. 3.8e-63;		Matches 577; Conservative 0; Mismatches 297; Indels 0; Gaps 0;	
QY	3	GGACATCGACACGTCAGCAGATCCGTTGGAGGACCTCACTGAGCCCGAGTGGGAGGACCT	62						
Db	1353	GGACATCGACACGACCTGGAGGCCCCGCTGTACTCACCCCGAGGGCTGGTCCCTCTT	1412						
QY	63	GACCCAGCAGTACTACTCCCTCGTTTCATGCGCATGCTTTCATCTCCCAATTCAAGAAATTA	122						
Db	1413	CCTCCAGCGCTACTACCAAGTGTCTCCAGAGGGGCGAGACTCAGGCACCTCGACACTCA	1472						
QY	123	CTTTTCGAGTCCGACGCTCTGCTGAATAGATCACTCTGTGTGAACCTCTCAGACGGACAT	182						
Db	1473	GGTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGTCAGCGCGGTGTATCCCGCAGATAGACAT	1532						
QY	183	TGACGGGCTCCGGAACATCTGGATTATAAGCCGCGCGCAAGTCCCGGGGCGGACACAT	242						
Db	1533	GGAAGGGGATCGCAACATCTGATCTGTGAAGCCAGGACCAAGTCCCGTGGACGAGGCAT	1592						
QY	243	AGTGTGATCGACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCACACCACTCTTTTC	302						
Db	1593	TATGTGATGACCACTTGAGGAGATCTGTGAAGCTGTGNAACGGCAACCCCGTGTGAT	1652						
QY	303	CAGGACAAACAGTGGTGTCTCCAGAAATACATCGAGACCCCGCTGCTCATCTGTGACAC	362						
Db	1653	GAAGGACGGCAAGTGGTGTGTCAGAAATATATTGAGCGGCCCTCTCATCTTTGGCAC	1712						
QY	363	CAAGTTGATCATGACAGAGTGGTCTCTGTCACGGACTGGAACCCCTGACCATCTGGTT	422						
Db	1713	CAAGTTTTCATCTCAGACAGTGGTCTCTGGTAACTGACTGGAACCCCACTTACCGTGTGGTT	1772						
QY	423	CTCAAGGAGAGTTACTTTCGGTCTTCAACTCAGCGCTTCTCCTGTCAGCAAGCTGGACAG	482						
Db	1773	CTACCGGCAGACTATATCGCTTTTCCACGACGCCCTTCTCCTGAGAACCTGGACAA	1832						
QY	483	CGCCATCCACTGTGCAACACCGCTCCAGAAATACCTGGAAGATGATGGGCGCGCAG	542						
Db	1833	CTCAGTGCACCTGTGCAACAACTCCATCCAGAGCACCTCGAGAACTCATGCGCCACGCA	1892						
QY	543	CCCCCTGCTCCCGCACACAACTGTGGACCGACGACAGCTTCAGAGTACCTGACGG	602						
Db	1893	TCCACTGCTTCCCGCACAACTATGTGTAGCTAGCAGAGTTCCAGGCCACCTCGCAGGA	1952						
QY	603	CCAGGCGCGTGGCGCGTGTGGGCGCGCTCTACCTACCGTCCATGAAGAAGGCCATCGC	662						
Db	1953	GATGGTGGCCCCAAATGCTTGTGTCACCATCATGTGCTGGCATGAGGATGCTGTGAT	2012						
QY	663	CCACGCGATGAAGTGGCCCGCAGGACCAACGTTGGAGCTTCGCAAGACAGCTTTGAGCTCTA	722						
Db	2013	CCACGCACTTCAGACCTCCCGAGGACCGTGCATGTCCGAGGCCAGCTTTGAGCTCTA	2072						
QY	723	CGGGCTGACTTGTCTTGGGAGGACTTCAGGCCCTTGCTGATCGAGATCAATTCAG	782						
Db	2073	TGCGCTGACTTGTGTTCGGGGTGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAG	2132						
QY	783	CCCACCATCAACCCGTCACGCGGTGTCAGGCCCGCTGCTGTGTCACAGCTGCAGGAGGA	842						
Db	2133	CCCACATGGGACCTTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTCAAGCTGA	2192						
QY	843	CACCATCAAGGTGGCGCTGGACCGCAGCTGTGAC	876						
Db	2193	CACCTTGGCGTGTCTATTGACCGGAGCTGGAC	2226						

RESULT 10
AK097236
LOCUS
DEFINITION
ACCESSION

AK097236 2848 bp mRNA linear PRI 30-JAN-2004
Homo sapiens cDNA FLJ39917 f1s, clone SPLEN2019405, highly similar
to Homo sapiens HOTT1 protein mRNA.
AK097236

VERSION
KEYWORDS
SOURCE
ORGANISM

AK097236.1 GI:21756925

Oligo capping; fis (full insert sequence).

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hiraoka, K., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Inose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Moniyama, H., Satoh, N., Takami, S.,
Teraashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takenoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shiraishi, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs

TITLE

JOURNAL
NAT. GENET. 36 (1), 40-45 (2004)
14702039

REFERENCE
AUTHORS

Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A.,
Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K.,
Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y.,
Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project

TITLE

REFERENCE

3 (bases 1 to 2848)

AUTHORS

Isogai, T. and Yamamoto, J.

TITLE

Direct Submission

JOURNAL

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES

Location/Qualifiers

1..2848

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="SPLEN2019405"

/tissue_type="spleen"

ACCESSION		AR339003	
VERSION		AR339003.1	
KEYWORDS		GI:33725860	
SOURCE		Unknown.	
ORGANISM		Unknown.	
REFERENCE		Unclassified.	
AUTHORS		1 (bases 1 to 3001)	
TITLE		Tang, Y.T., Zhou, P. and Drmanac, R.T.	
JOURNAL		Nucleic acids and polypeptides	
FEATURES		Patent: US 6569662-A 494 27-MAY-2003;	
source		Location/Qualifiers	
		1..3001	
		/organism="unknown"	
		/mol_type="genomic DNA"	
ORIGIN			
Query Match		44.0%; Score 398.8; DB 6; Length 3001;	
Best Local Similarity		66.3%; Pred. No. 3.8e-63;	
Matches 574; Conservative		0; Mismatches 292; Indels 0; Gaps 0;	
Qy	3	GGACATCGACAGCTGAGGAGCTGCGTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCT	62
Db	1363	GGACATCGACAGGACCTGGAGGCCCGCTGTACTCTACCCCGAGGGTGGTCCCTCTTT	1422
Qy	63	GACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATCAAGAAATTA	122
Db	1423	CCTCCAGCGCTACTACCAAGTGGTCCACGAAGGGGCAGAACTCAGGCACCTCGACACTCA	1482
Qy	123	CTTTTCGAGTGCACAGGCTCTGCTGGAATAGAAATCAGTCTGTGAACCTCTAGACGGACAT	182
Db	1483	GGTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGTCAGGCCGTGGTATCCCGAGATAGACAT	1542
Qy	183	TGACGGGCTCCGGAACTCTCGATTATAAAGCCCGCGCCAAAGTCCCGGGCCCGAGACAT	242
Db	1543	GGAAAGGGAATGCAACATCTCGATCGTGAAGCCAGAGCCAAAGTCCCGCGAGCGAGCAT	1602
Qy	243	AGTGTGCATGGAACCGTGTGGAGGAGATCTCTGAGCTGGCAGCTGCAGACCAACCTCTTTTC	302
Db	1603	CATGTGCATGGACCACTTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGAT	1662
Qy	303	CAGGAGCAACAAAGTGGGTGTCAGAAAGTACATCGAGACGCCGTGCTCATCTGTGACAC	362
Db	1663	GAAGGACGGCAAGTGGGTGGTGCAGAAAGTATATTTGAGCGGCCCTCTCTCATCTTTGGCAC	1722
Qy	363	CAAGTTTCGACATCGACAGTGTCTCTGTCACCGACTTGAACCCCTCTGACCACTCTGGTT	422
Db	1723	CAAGTTTGACCTCAGACAGTGTCTCTGCTGAACCTGCTGTAACCTGTAACCCACTTACCGTGGTT	1782
Qy	423	CTACAAGGAGATTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAG	482
Db	1783	CTACCGGACAGCTATATCCGCTTTCCACGACGCCCTTCTCTCTGAAGAACCTGGACAA	1842
Qy	483	CGCCATTCACCTGTGCAACAAAGCGCTGCAGAAATGATCTGAAGAAATGATGTGGGCCGAC	542
Db	1843	CTCAGTGCACTGTGCAACAACTCCATCCAGAAACACTGAGAAATCTCATGCCATCGGCA	1902
Qy	543	CCCGCTGTGCGCGCACACACATGTCGACCAACAGTTCAGAGTTCAGAGTACCTTGGACCG	602
Db	1903	TCCACTGTCTTCGCGCAGCAACATGTGGTCTAGCGTTCAGAGGTTCCAGGCCCTCTGGACGA	1962
Qy	603	CCAGGCGCTGCGCGCGTGTGGGCGAGCGTTCATCTACCCGTCATGAAGAGGCGCATTCGC	662
Db	1963	GATGGGTGCCCCAATGTTGGTCCACCATCATCTGCTGGCATGAAGGATGCTGTGAT	2022
Qy	663	CCACGCCATGAAGTGGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTCTA	722
Db	2023	CCACGCACCTTCAGACCTCCAGGACACCGTGCAGTGTGGAAAGGCCAGCTTTGAGCTCTA	2082
Qy	723	CGGGCTGACTTCGTTCTTGGAGGAGCTTCAGGCCCTGGCTGATCAGATCAATTCACG	782
Db	2083	TGGCGCTGACTTCGTTCTTGGAGGAGCTTCAGGCCCTGGCTGATCAGATCAACGCGCAG	2142
Qy	783	CCCCACCAATGACCCCGTCCACGCGCGTCAACGCGCGTGTGTGTCACAGGTGTGAGGAGGA	842
RESULT 11			
AR339003			
LOCUS			
Sequence 494 from patent US 6569662.			
DEFINITION			
AR339003			
3001 bp			
DNA			
linear			
PAT 17-AUG-2003			

RESULT 11
AR339003
LOCUS AR339003 3001 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 494 from patent US 6569662.

Db 2143 CCCCACCATGGCAGCCCTCCACAGCAGTCACTGCCCGCTCTGTGCTGGCGTGAAGCTGA 2202
 Qy 843 CACCATTCAAGTGGCCGTGGACCGCA 868
 Db 2203 CACCTGGCGTGTCAATTGACCGGA 2228

RESULT 12
 AF078842
 LOCUS 1684 bp mRNA linear PRI 09-JAN-2000
 DEFINITION Homo sapiens HOTT1 protein mRNA, complete cds.
 ACCESSION AF078842
 VERSION AF078842.1 GI:6683744
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1684)
 AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
 Zhou, S., Liu, M. and He, F.
 TITLE Functional prediction of the coding sequences of 50 new genes
 deduced by analysis of cDNA clones from human fetal liver
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1684)
 AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
 Zhou, S., Liu, M. and He, F.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-1998) Dept. of Experimental Hematology, Beijing
 Institute of Radiation Medicine, 27 Taiping RD, Beijing 100850,
 P.R.China

FEATURES
 Location/Qualifiers
 1..1684
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HQ0207"
 /tissue type="liver"
 /dev stage="fetus"
 1..398
 399..1277
 /notes="hgp0207; similar to bovine and pig tubulin-tyrosine
 ligase (TTL); Swiss-Prot Accession Numbers P38584 and
 P38160"
 /codon_start=1
 /evidence=not_experimental
 /product="HOTT1 protein"
 /protein_id="AAF23353.1"
 /db_xref="GI:6683745"
 /translation="MEGRNFWIVKPGKSRGIMCMHDLBMLKLVNPNVPMKDG
 KWVQKIERPLHFCGFKDLRQFELVTWNPFLTVMFYRDSYIRFSTQPSLKNLNS
 VHLNNSIQKHLNSCHRHFLFPDNNMSQRFQALHNGAENANSTIIVFGMKDAV
 IHALQTSQTVQCKASPELLYGADFPVGEDFPQLIEINASTPMASPTAVTAKLCAGV
 QADTLRVVIDMLDRNCDTGAFELIYKQPVTPSPATSPRPSCLLPMPYSDTRARSSDS
 TASWMLRPCRPOARP"
 1278..1684

5'UTR
 CDS

3'UTR
 ORIGIN

Query Match 41.6%; Score 376.8; DB 9; Length 1684;
 Best Local Similarity 69.3%; Pred. No. 4.3e-59;
 Matches 513; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

Qy 129 GCAGTGCAGGCTCTGCTGATAGAAATCAGTCTGTGAACCCCTCAGCAGGACATTGACGG 198
 Db 347 GCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCGTGTATCCCCAGATACATGGAAGG 406
 Qy 189 GCTCCGGAACATCTGGATTATAAGCCCGCGGCCCAAGTCCCGGGCCGAGACATAGTG 248
 Db 407 GGATCGCACTTCTGGATCTGGAAGCAGGAGCCAGTCCCGGGACAGGCATCATGTG 466
 Qy 249 CATGGACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACACACCTCTTCCAGGGA 308

TITLE Schnersch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2238257
MEDLINE 12477932
REFERENCE 2 (bases 1 to 1897)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 16 Row: g Column: 10.
Location/Qualifiers

FEATURES
source

1. .1897
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:11856 IMAGE:3597662"
/tissue_type="mammary tumor. C3(1)-Tag model. Infiltrating
ductal Carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .1897
/gene="4833441J24R1k"
/db_xref="LocusID:101100"
/db_xref="MGI:2141418"
590. .1390
/gene="4833441J24R1k"
/codon_start=1
/product="4833441J24R1k protein"
/protein_id="AAH06830.1"
/db_xref="GI:13905090"
/db_xref="LocusID:101100"
/db_xref="MGI:2141418"
/translation="MEGDQNIWVKPGAKSRGIMCMNRLDMLKLVDCNPMKMDG
KVIQVYIRPFLIFGTFDLQWFLVTDWNPFLVWFYRDSVIRFSTOPFSKLNLDNS
VHLNNSIQRLHLEASCHPMLPDDNMWSSQFOAHLQEVDAKPAWSSVIVPQMAAV
IHALQTSQNVQCRASFELYGADFVFGEDFQFLIEINASTMAPSTAVARLCAGV
QADTLRVVDRDLRSDCTGAFELIYKQVRCSSSTSHKQHPGPGWCET"
590. .1270
/gene="4833441J24R1k"
/note="TTL; Region: Tubulin-tyrosine ligase family.
Tubulins and microtubules are subjected to several
post-translational modifications of which the reversible
detyrosination/tyrosination of the carboxy-terminal end of
most alpha-tubulins has been extensively analysed. This
modification cycle involves a specific carboxypeptidase
and the activity of the tubulin-tyrosine ligase (TTL). The
true physiological function of TTL has so far not been
established. Tubulin-tyrosine ligase (TTL) catalyses the
ATP-dependent post-translational addition of a tyrosine to
the carboxy terminal end of detyrosinated alpha-tubulin.

misc_feature

In normally cycling cells, the tyrosinated form of tubulin
predominates. However, in breast cancer cells, the
detyrosinated form frequently predominates, with a
correlation to tumour aggressiveness. On the other hand,
3-nitrotyrosine has been shown to be incorporated, by TTL,
into the carboxy terminal end of detyrosinated
alpha-tubulin. This reaction is not reversible by the
carboxypeptidase enzyme. Cells cultured in 3-nitrotyrosine
rich medium showed evidence of altered microtubule
structure and function, including altered cell morphology,
epithelial barrier dysfunction, and apoptosis"
/db_xref="CDD:pfam03133"

ORIGIN

Query Match 40.4%; Score 366.4; DB 10; Length 1897;
Best Local Similarity 69.3%; Pred. No. 3.4e-57;
Matches 499; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
QY 163 GTGAACCTTCAGACGGACATTGACCGGCTCCGGAACATCTGGATTATAAAGCCCGCGCC 222
DB 572 GTGTACCCAGTTAGACATCGAGGGGATCAGACATCTGGATCGTGAAGCCCGAGCC 631
QY 223 AAGTCCCGGGCGGAGACATAGTGTGCATGACCGGTGTGAGAGAGATCTCGAGCTGGCA 282
DB 632 AAGTCCCGAGCGGAGGATTTATGTCATGAACCGCTGGATGAGATGCTGAAGCTGGT 691
QY 283 GCTGCAGACACCTCTTTCCAGGAGACAAAGTGGTGTCCAGAACTACATCGAGACG 342
DB 692 GACTGCAACCCATGCTCATGAAGATGCAAGTGTGTCGAGAAATGATCAATTGACGG 751
QY 343 CGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAAGACTGG 402
DB 752 CCCCTGCTTATCTTCGGCACCAAGTTTGACCTTGAGACAGTGGTTCCTGCTGACTGG 811
QY 403 AACCCCTGACCATCTGTTCTACAAGGAGAGTACTTCCGTTTCTCAACTCAGCCCTTC 462
DB 812 AACCCACTCACGCTGGTGTCTACCGAGACAGTCACTTCGTTTCTCCACACAGCCCTTC 871
QY 463 TCCCTGGACAAGCTGCAGACGGCCATCCACCTGTGCAACAAACCGCTCCAGAACTACCTG 522
DB 872 TCCCTGAAGAACCTGACAACTCTGTGCACTGTGTAACTCTCCATCCAGAGACTTG 931
QY 523 AAGAAATGATGTGGCGCGCAGCCCTCTGTGCTCCCGCACACAAATGTGGACACACAG 582
DB 932 GAGGCTTCCTGTACCGGCACCGATGCTGCCCCAGATAACATGTGGTCCAGCCAGAG 991
QY 583 TTCCAGGAGTACTCTGACGGCCAGGGCCGTGGCGCGGTGGGGGAGCGTCACTTACCCG 642
DB 992 TTTGAGGCCCACTTTCAGGAGGTAGATGCCCCAAAGGCCGTGGTCCAGCGTCACTCGTGCCA 1051
QY 643 TCCATGAAGAAGGCCATCGCCACCGCATGAAGGTGGCCCGAGACCCACGTGGAGCCTCGC 702
DB 1052 GGCATGAAGGTGCTGTGATCCAGCCCTGCGAGACCTCCAGACACACGTGAGTGCCTGG 1111
QY 703 AAGAAACAGCTTTGAGCTTACGGGGCTGACTTCGTCTTCTTGGAGGAGACTTCAGGCCCTCG 762
DB 1112 AAGGCCAGCTTTGAGCTTATGGGGCAGACTTTGTGTGTGGGGAAGACTTCCAGCCCTCG 1171
QY 763 CTGATGAGATCAATTCAGCCCGCCACCATGACCCCGTCCAGCCGGTTCAGGCCCGAGCTG 822
DB 1172 TTGATGAATCAATGCCAGGCCCACTTCCAGCCCTTCCAGCGGTGTCTACGCCGCCCTC 1231
QY 823 TGTGCAGAGGTGCAGGAGACACCATCAAGGTGGCGGTGGACCGGAGCTGTGACATCGGC 882
DB 1232 TGTGCGGTGTGAGGACGATACCTTCGCTGTGCTCATCGACCGGCCCTTGACCGCAGC 1291

RESULT 14

AC119959
LOCUS Mus musculus chromosome 15, clone RP24-467H19, complete sequence.
AC119959
DEFINITION AC119959
ACCESSION AC119959
VERSION AC119959.8 GI:50540805

KEYWORDS
SOURCE
ORGANISM

HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 163783)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus chromosome 15, clone RP24-467H19

TITLE
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 163783)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Submitted (02-MAY-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

3 (bases 1 to 163783)
Birren, B., Nusbaum, C., Lander, E., Abouellail, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J.,
Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
McClean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (26-MAY-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 163783)
Birren, B., Nusbaum, C., Lander, E., Abouellail, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J.,
Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
McClean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,

Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL

Direct Submission
Submitted (23-JUL-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Jul 23, 2004 this sequence version replaced gi:47679219.
All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L25717

Center clone name: 467_H_19

FEATURES

source

Location/Qualifiers

1..163783

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP24-467H19"

/clone_lib="RP24-467H19"

1..4

/note="clone boundary"

clone_end:SP6

site:MboI

complement(169..204)

/rpt_family="ORR1B1"

818..994

/rpt_family="ORR1D"

1312..1527

/rpt_family="B3"

1573..1645

/rpt_family="ID3"

1665..1793

/rpt_family="B1F"

1794..1959

/rpt_family="B3"

complement(2256..2658)

/rpt_family="L1"

complement(2659..2833)

/rpt_family="B2_Mm2"

complement(2834..3302)

/rpt_family="L1"

complement(3311..3448)

/rpt_family="B1_MM"

complement(3449..4218)

/rpt_family="L1"

4214..4557

/rpt_family="Lx6"

5120..5279

/rpt_family="CT-rich"

complement(5297..5423)

/rpt_family="B1F"

complement(5451..5524)

/rpt_family="PB1D9"

6016..6494

/rpt_family="RLTR11A"

complement(6678..6797)

/rpt_family="B1_MM"

complement(6810..6945)

repeat_region	/rpt_family="B1_MM" 7008...7148	
repeat_region	/rpt_family="B1_MM" complement(7763...7941)	
repeat_region	/rpt_family="B3" complement(7960...8105)	
repeat_region	/rpt_family="B1_MM" complement(8121...8300)	
repeat_region	/rpt_family="B3A" 8486...8625	
repeat_region	/rpt_family="B4A" 8618...8788	
repeat_region	/rpt_family="B1_MM" 8804...9102	
repeat_region	/rpt_family="L1MB2" 9162...9217	
repeat_region	/rpt_family="PB1D7" complement(9339...9460)	
repeat_region	/rpt_family="B3" complement(9598...9748)	
repeat_region	/rpt_family="B3" complement(9755...12502)	
repeat_region	/rpt_family="L1_MM" 12503...12992	
repeat_region	/rpt_family="IAPLTR2_MM" 12994...13717	
repeat_region	/rpt_family="IAPLTR2_MM-int" 13718...13836	
repeat_region	/rpt_family="IAPLTR2_MM-int" 13832...14083	
repeat_region	/rpt_family="IAPLTR2_MM-int" 14083...15904	
repeat_region	/rpt_family="IAPLTR2_MM-int" 15905...16394	
repeat_region	/rpt_family="IAPLTR2_MM" complement(16395...20012)	
repeat_region	/rpt_family="L1" complement(20013...20157)	
repeat_region	/rpt_family="B1_MM" 20192...20290	
repeat_region	/rpt_family="B2_Mm2" 20410...20553	
repeat_region	/rpt_family="B1_MM" 20586...20777	
repeat_region	/rpt_family="B2_Mm2" 21012...21082	
repeat_region	/rpt_family="CCAA)n" 21192...21294	
repeat_region	/rpt_family="PB1D10" 21301...21342	
Query Match	31.1%; Score 281.6; DB 10; Length 163783;	
Best Local Similarity	78.7%; Pred.No. 6.4e-42;	
Matches 354; Conservative 0; Mismatches 84; Indels 12; Gaps 1;		
QY 469	GACAAGCTGACAGCGCCATCCACCTGTGTGCAACAGCGGTCCAGAACTGACCTGGAAGAT 528	
Db 136973	GACATTGCCACAGCGCCATCCACCTGTGTACAACTCCATCCAGAGCGCTCAAAAT 137032	
QY 529	GATGTGGCGGACGCCCCCTGTGTGCCGACACAAATGTGACACGACACAGCTTCCAG 588	
Db 137033	GATAAAGAGCGGAGTCCGCTGTCTACCTTGTCTAATGTGACACGACTCGTTTCCAG 137092	
QY 589	GAGTACCTCAGCGCCAGCGCGGTGTGGGCGAGCGGTCTATCTACCGTCCATG 648	
Db 137093	GAGTACCTCAGAGAGGGCGGAGGAGAACTGTGGGTAGATCATCTACCGCTCTATG 137152	
QY 649	AAAGAGCCATCCGCCACGCGCATGAAGTGGGCCAGGACCACTGGAGCGCTCGCAAGAAC 708	
Db 137153	AAAAGAGCTGTACCAATGCCATGAGGGTAGCCAGGACCACTAGAACCGCGTAAAGAC 137212	
QY 709	AGCTTTGAGCTACGGGCTGATCTTCGTCCTTGGGAGGAGCTTACGGCGCTGGCTGATC 768	

Db 137213	AGCTTCGAGCTCTACGGAGCCGACTTCATCTCTGGGGCGAGACTTCAAGCATGGCTTATC 137272	
QY 769	GAGATCAATTCCAGCCGCCACCATGACACCGGTCCACGCCGTCACGCCAGCTGTGTGCA 828	
Db 137273	GAGATCAACTCCAGCCGCCACCATGACACCGTCCATCCGTCCAGCTCAGCTCTGTGCC 137332	
QY 829	CAGGTGCAGGAGGACACCATCAAGTGGCGC-----TGGACGCGAGCTGTGAC 876	
Db 137333	CAGGTGCAGGAGGACACCATCAAGTGGTGGTGGATCGCAAACTGGACCGAAACTGTGAC 137392	
QY 877	ATCGGCAACTTCGAGCTCTCTGTGGGCGAG 906	
Db 137393	ATCGGCAACTTTCGAGCTCTCTGTGGGCGAG 137422	
RESULT 15		
AC117700	201420 bp DNA linear HTG 26-SEP-2004	
LOCUS	Mus musculus chromosome 15 clone RP23-455J20 map 15, *** SEQUENCING	
DEFINITION	IN PROGRESS ***, 3 unordered pieces.	
ACCESSION	AC117700	
VERSION	AC117700.17 GI:52694698	
KEYWORDS	HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	1 (bases 1 to 201420)	
AUTHORS	Birren,B., Nusbaum,C. and Lander,E.	
TITLE	Mus musculus chromosome 15, clone RP23-455J20	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 201420)	
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczkzy,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Traversa,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
REFERENCE	3 (bases 1 to 201420)	
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., McEwan,P., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,	

Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thonann, N.,
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodor, J.,
Tropham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (26-SEP-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 26, 2004 this sequence version replaced gi:52077711.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L24098

Center clone name: 455_J_20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 82149: contig of 82149 bp in length
* 82150 82249: gap of unknown length
* 98003: contig of 15754 bp in length
* 98004 98103: gap of unknown length
* 98104 201420: contig of 103317 bp in length.

FEATURES

source

1. .201420
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-455J20"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match 31.1%; Score 281.6; DB 2; Length 201420;
Best Local Similarity 78.7%; Pred. No. 6.2e-42;
Matches 354; Conservative 0; Mismatches 84; Indels 12; Gaps 1;

QY 469 GACAAGCTGACAGCGCCATCCACCTGTGTGCAACAACGCCGTCAGAGTACCTGAAGAAT 528
Db |||||
QY 4228 GACATTGCCACAGCGCCATCCACCTGTGTAAACAATCCATCCAGAGCGCCTCAAAAAT 4287
Db |||||
QY 529 GATGTGGCGCAGCGCCCGCTGTGCCGACACAACATGTGGACGACGACGAGTTCCAG 588
Db |||||
QY 4288 GATAAAGAGCGAGTCGGCTGCTACCTTGTATAACATGTGGACGACACTCGTTTCCAG 4347
Db |||||
QY 589 GAGTACCTGACGCGCCAGCGCCGCTGTGGGGGACGCTCATCTACCGGTCCATG 648
Db |||||
QY 4348 GAGTACCTGACGAGGGGCGGAGGAGAACGTGGGTAGCATCTACCCGCTCTATG 4407
Db |||||
QY 649 AAGAAGGCCATCGCCACGCCATGAAGGTGGCCGACGACCACTGGAGCCTCGCAAGAAC 708
Db |||||
QY 4408 AAAAGAGCTGTCAACATGCGATGAGGTTAGCCGACGACCACTAGAACGCCGTAAGAAC 4467
Db |||||
QY 709 AGCTTTGAGCTCTACGGGGCTGACTTCTCTTGGGAGGGAATTCAGGCCCTGGCTGATC 768
Db |||||
QY 4468 AGCTTCGAGCTCTACGGAGCGGACTTCTATCTCGGGCGGAGACTTCAAGCCATGGCTTATC 4527
Db |||||
QY 769 GAGATCAATTCAGGCCCGACCGCTGACGCGGCTCACGGCCAGCTGTGTGCA 828
Db |||||
QY 4528 GAGATCAATTCAGGCCCGACCGCTGACGCGGCTCACGCTCAGCTCAGCTCTGTGCC 4587
Db |||||

QY 829 CAGGTGCAGGAGGACACCATCAAGTGGCGG-----TGGACCGCAGCTGTGAC 876
Db |||||
QY 4588 CAGGTGCAGGAGGACACCATCAAGTGGTGGATCGCAAACTGGACCGAAACTGTGAC 4647
Db |||||
QY 877 ATCGGCAACTTTCGAGCTCCTGTGGAGGCGAG 906
Db |||||
QY 4648 ATCGGCAACTTTGAGCTTCTGTGGCGGCGAG 4677
Db |||||

Search completed: September 25, 2005, 11:26:56

Job time : 4207.65 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 23:48:14 ; Search time 1252.07 Seconds
(without alignments)
8666.982 Million cell updates/sec

Title: US-10-615-659-1_COPY_153_1775
Perfect score: 1623
Sequence: 1 agggatccagcatctcaaa.....cgccggaggaaagatttca 1623

Scoring table: IDENTITY NUC
Gapop 10.0 , Capext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17:	/cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18:	/cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19:	/cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20:	/cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21:	/cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22:	/cgn2_6/ptodata/2/pubpna/US10J_NEW_PUB.seq:*
23:	/cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24:	/cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1623	100.0	1838	19	US-10-615-659-1
2	1623	100.0	1838	19	US-10-615-977-1
3	1619.8	99.8	3554	19	US-10-615-659-12
4	1619.8	99.8	3554	19	US-10-615-977-12
5	1505.8	92.8	1859	19	US-10-615-659-10
6	1505.8	92.8	1859	19	US-10-615-977-10
7	1475.8	90.9	3465	19	US-10-615-659-11
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 12, Appl
					Sequence 10, Appl
					Sequence 11, Appl

8	1475.8	90.9	3465	19	US-10-635-977-11	Sequence 11, Appli
9	1474.2	90.8	1939	19	US-10-615-659-9	Sequence 9, Appli
10	1474.2	90.8	1939	19	US-10-635-977-9	Sequence 9, Appli
11	726	44.7	726	19	US-10-615-659-3	Sequence 3, Appli
12	726	44.7	726	19	US-10-635-977-3	Sequence 3, Appli
c 13	501	30.9	101270	20	US-10-723-860-631	Sequence 631, App
c 14	432.8	26.7	2380	18	US-10-275-595A-39	Sequence 39, Appli
c 15	427	26.3	190	16	US-10-029-386-24894	Sequence 24894, A
16	421.4	26.0	2538	15	US-10-102-524-1765	Sequence 1765, Ap
17	421.4	26.0	2538	21	US-10-956-157-2016	Sequence 2016, Ap
18	421.4	26.0	2538	22	US-10-756-149-1371	Sequence 1971, Ap
19	421.4	26.0	3001	15	US-10-037-270-494	Sequence 494, App
20	421.4	26.0	3001	17	US-10-117-722-494	Sequence 494, App
21	418	25.8	2848	17	US-10-108-260A-1766	Sequence 1766, Ap
22	413	25.4	2611	9	US-09-925-298-199	Sequence 199, App
23	413	25.4	2611	14	US-10-102-806-199	Sequence 199, App
c 24	387.6	23.9	526	13	US-10-027-632-135927	Sequence 135927,
c 25	387.6	23.9	526	13	US-10-027-632-135928	Sequence 135928,
c 26	387.6	23.9	526	17	US-10-027-632-135927	Sequence 135927,
c 27	387.6	23.9	526	17	US-10-027-632-135928	Sequence 135928,
28	381.8	23.5	1400	21	US-10-956-157-9356	Sequence 9356, Ap
29	381.8	23.5	1673	21	US-10-956-157-4121	Sequence 4121, Ap
30	319.8	19.7	1400	21	US-10-956-157-7251	Sequence 7251, Ap
31	196	12.1	755	13	US-10-027-632-135929	Sequence 135929,
32	186	12.1	755	13	US-10-027-632-135930	Sequence 135930,
33	186	12.1	755	17	US-10-027-632-135929	Sequence 135929,
34	196	12.1	755	17	US-10-027-632-135930	Sequence 135930,
35	190.6	11.7	1958	21	US-10-956-157-4903	Sequence 4903, Ap
36	166	10.2	1728	22	US-10-450-763-5962	Sequence 5962, Ap
37	164.4	10.1	2250	22	US-10-450-763-5187	Sequence 5187, Ap
38	163.4	10.1	418	9	US-09-983-965-2032	Sequence 2032, Ap
39	131.8	8.1	492	10	US-09-918-995-35253	Sequence 35253, A
c 40	130	8.0	525	16	US-10-029-386-11174	Sequence 11174, A
41	122.6	7.6	436	17	US-10-242-535A-7660	Sequence 7660, Ap
42	122.6	7.6	436	18	US-10-085-783A-7660	Sequence 7660, Ap
43	120.4	7.4	2241	19	US-10-615-659-27	Sequence 27, Appl
44	120.4	7.4	2241	19	US-10-635-977-27	Sequence 27, Appl
45	113.4	7.0	279	18	US-10-424-599-119452	Sequence 119452,

ALIGNMENTS

RESULT 1
US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; TYPE: DNA
; LENGTH: 1838
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Query Match 100.0%; Score 1623; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCATCCAGCATCTCAAGTGGTGGTCAGCCACAGAGCTGCAGCAGCAGCAGA 60
|||||

Db 153 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCTGCAGGAGCAGCAGA 212
QY 61 AGCAAGCCAGGACACAGAGGAGAGGCGCGGAGCAGCGACCTGTAGCAGCAGGCAAGAT 120
Db 213 AGCAAGCCAGGACACAGAGGAGAGGCGCGGAGCAGCGACCTGTAGCAGCAGGCAAGAT 272
QY 121 GCTGAAAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGACGCTTGTGTGACATCGCGTGC 180
Db 273 GCTGAAAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGACGCTTGTGTGACATCGCGTGC 332
QY 181 AAGGTGTGCAGGCTTACCTGGGCGAGCTGGAGCATGAGGACATCGACAGCTCAGCAGAT 240
Db 333 AAGGTGTGCAGGCTTACCTGGGCGAGCTGGAGCATGAGGACATCGACAGCTCAGCAGAT 392
QY 241 GCCGTGGAGGACCTCACTCAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
Db 393 GCCGTGGAGGACCTCACTCAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 452
QY 301 GTTCATGGCGATGCTTTTCATCTCCAATTCAGAAATTAATTTTCGAGTGCACAGGCTGTG 360
Db 453 GTTCATGGCGATGCTTTTCATCTCCAATTCAGAAATTAATTTTCGAGTGCACAGGCTGTG 512
QY 361 CTGCAATAGAAATCAGCTGTGAACCTCAGACGACATTTGACGGGCTCCGGACATCTGG 420
Db 513 CTGCAATAGAAATCAGCTGTGAACCTCAGACGACATTTGACGGGCTCCGGACATCTGG 572
QY 421 ATTATAAAGCCCGGCGCAAGTCCCGGGGCCGAGACATAGTGTGATGACCGTGTGGAG 480
Db 573 ATTATAAAGCCCGGCGCAAGTCCCGGGGCCGAGACATAGTGTGATGACCGTGTGGAG 632
QY 481 GAGATCCTGGAGTGGCAGCTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTC 540
Db 633 GAGATCCTGGAGTGGCAGCTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTC 692
QY 541 CAGAAATGATCAGACGCGGCTGTCTCATCTGTGACACCAAGTTGACATCAGACAGTGG 600
Db 693 CAGAAATGATCAGACGCGGCTGTCTCATCTGTGACACCAAGTTGACATCAGACAGTGG 752
QY 601 TTCTCTCGTCACGAGCTGGAACCCCTGACCATCTGGTCTCAAGAGGAGATTACTTGGG 660
Db 753 TTCTCTCGTCACGAGCTGGAACCCCTGACCATCTGGTCTCAAGAGGAGATTACTTGGG 812
QY 661 TTCTCAACTCAGGCTTCTCCTTGACAAGCTGGACAGCGCCATCCACTGTGCAACAC 720
Db 813 TTCTCAACTCAGGCTTCTCCTTGACAAGCTGGACAGCGCCATCCACTGTGCAACAC 872
QY 721 GCGGTCAGAACTACTGGAAGATGATGTGGCGCGAGCCCTGCTGCGCCGACACAC 780
Db 873 GCGGTCAGAACTACTGGAAGATGATGTGGCGCGAGCCCTGCTGCGCCGACACAC 932
QY 781 ATGTGGACACGACCAAGGTTCCAGAGTACTCTGACGATCTGGTCTCAAGAGGAGTTACTTGGG 840
Db 933 ATGTGGACACGACCAAGGTTCCAGAGTACTCTGACGATCTGGTCTCAAGAGGAGTTACTTGGG 992
QY 841 GCGAGGTCATCTACCCGTCCATGAAGAGGCCATCGGCCACGCCATGAAGTGGCCCCAG 900
Db 993 GCGAGGTCATCTACCCGTCCATGAAGAGGCCATCGGCCACGCCATGAAGTGGCCCCAG 1052
QY 901 GACACGTGGAGCCTCGAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG 960
Db 1053 GACACGTGGAGCCTCGAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG 1112
QY 961 AGGGAATTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCCGTCCAG 1020
Db 1113 AGGGAATTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCCGTCCAG 1172
QY 1021 CCGGTACGCGCCAGCTGTGTGCACAGGTGCAGGAGCACCATCAAGTGGCCGTGGAC 1080
Db 1173 CCGGTACGCGCCAGCTGTGTGCACAGGTGCAGGAGCACCATCAAGTGGCCGTGGAC 1232
QY 1081 CGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGGAGGACCGGCTGGTTCAGCGGCC 1140
Db 1233 CGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGGAGGACCGGCTGGTTCAGCGGCC 1292

QY 1141 CCATTTACGGGCTCGACCTCTGCGTGGCGGCGTCACTGTGAGGAGCCAGGAGGCAG 1200
Db 1293 CCATTTACGGGCTCGACCTCTGCGTGGCGGCGTCACTGTGAGGAGCCAGGAGGCAG 1352
QY 1201 GTGTGCGCGCTCTGCAACCTCAAGGCTCGGCTCGCTTGTGTGACGCGAGCGCTGAAG 1260
Db 1353 GTGTGCGCGCTCTGCAACCTCAAGGCTCGGCTCGCTTGTGTGACGCGAGCGCTGAAG 1412
QY 1261 GCACGGGGCCCTCGGCGCATGCAGACCTGCCAGGAGCCCCATCACAGCTCTCCAG 1320
Db 1413 GCACGGGGCCCTCGGCGCATGCAGACCTGCCAGGAGCCCCATCACAGCTCTCCAG 1472
QY 1321 CGGGAATTTGGGACTGAAGGAGAGAGGGGCTCCCTCTGGCCCTTGTGTGCACCCCTTAAAG 1380
Db 1473 CGGGAATTTGGGACTGAAGGAGAGAGGGGCTCCCTCTGGCCCTTGTGTGCACCCCTTAAAG 1532
QY 1381 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGAAGGTG 1440
Db 1533 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGAAGGTG 1592
QY 1441 GAGCTCCCGGCTGCGCTGTGCGCACGTTGGACAGTCAAGGCCCCAAAACACCGGTGCCCC 1500
Db 1593 GAGCTCCCGGCTGCGCTGTGCGCACGTTGGACAGTCAAGGCCCCAAAACACCGGTGCCCC 1652
QY 1501 GTAGCCCAAGCCCGCAAAAGCTGGGATCCAAACCAAGTAAATGCCGACCCCGCTGGAGCCT 1560
Db 1653 GTAGCCCAAGCCCGCAAAAGCTGGGATCCAAACCAAGTAAATGCCGACCCCGCTGGAGCCT 1712
QY 1561 GTGTGCGGGGCTGAAGACAGACAGAGGCGCGCTGCGTCCGCGCCGAGGAGAAAGT 1620
Db 1713 GTGTGCGGGGCTGAAGACAGACAGAGGCGCGCTGCGTCCGCGCCGAGGAGAAAGT 1772
QY 1621 TCA 1623
Db 1773 TCA 1775

RESULT 2

US-10-635-977-1
; Sequence 1, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-635-977-1

Query Match 100.0%; Score 1623; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCTGCAGGAGCAGCAGA 60
Db 153 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCTGCAGGAGCAGCAGA 212
QY 61 AGCAAGCCAGGACACAGAGGAGAGGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGAT 120

213 AGCAAGCCAGGAGCCAGAGGGAGGAGCCGGGAGCAGCAGCTGAGCAGCAGGACAGAT 272
121 GCTGAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 180
273 GCTGAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 332
181 AAGGTGTCAGAGCCCTACCTGAGGGAGCTGAGGAGTGGAGACATCGACAGTCTCAGAGAT 240
333 AAGGTGTCAGAGCCCTACCTGAGGGAGCTGAGGAGTGGAGACATCGACAGTCTCAGAGAT 392
241 GCGGTGAGGACCTCAGTGAAGGCGAGTGGAGGAGCTGACCCAGCAGTACTACTCCCTC 300
393 GCGGTGAGGACCTCAGTGAAGGCGAGTGGAGGAGCTGACCCAGCAGTACTACTCCCTC 452
301 GTTCATGGCGATGTTTCATCTCAATTCAGAAATTTACTTTTGGCAGTCCCAAGGCTCTG 360
453 GTTCATGGCGATGTTTCATCTCAATTCAGAAATTTACTTTTGGCAGTCCCAAGGCTCTG 512
361 CTGAATAGATCAGCTCTGTGAACCTTCAGACGAGACATTCACGGGCTCCCGAATCATCTGG 420
513 CTGAATAGATCAGCTCTGTGAACCTTCAGACGAGACATTCACGGGCTCCCGAATCATCTGG 572
421 ATTATTAAGCCCGGGCCAAAGTCCCGGGGCGGAGACATAGTGTGCATGGACCGTGTGGAG 480
573 ATTATTAAGCCCGGGGCAAGTCCCGGGGCGGAGACATAGTGTGCATGGACCGTGTGGAG 632
481 GAGATCTGAGCTGGCAGCTGACAGCCCTCTTTCCAGGGAGCAACAAGTGGGTGTC 540
633 GAGATCTGAGCTGGCAGCTGACAGCCCTCTTTCCAGGGAGCAACAAGTGGGTGTC 692
541 CAGAGTACATCAGAGCCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
693 CAGAGTACATCAGAGCCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 752
601 TTCTCTGTACGAGCTGGAACCCCTGCACCATCTGGTTCTACAAGGAGAGTACTTTCGCG 660
753 TTCTCTGTACGAGCTGGAACCCCTGCACCATCTGGTTCTACAAGGAGAGTACTTTCGCG 812
661 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 720
813 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 872
721 GCGGTGACAGAGTACCTGAAGAATGATGTGGGCGGAGCCCTCTGTCGCCGACACAAAC 780
873 GCGGTGACAGAGTACCTGAAGAATGATGTGGGCGGAGCCCTCTGTCGCCGACACAAAC 932
781 ATGTGGACACAGCAGGTTCCAGGAGTACCTGCAGCGCCAGGCGCGTGGCGCGTGTGG 840
933 ATGTGGACACAGCAGGTTCCAGGAGTACCTGCAGCGCCAGGCGCGTGGCGCGTGTGG 992
841 GGCAGCGTCATCTACCGCTCCATGAAGAAGGCCATCGCCACGCCATGAAGGTGGCCCGAG 900
993 GGCAGCGTCATCTACCGCTCCATGAAGAAGGCCATCGCCACGCCATGAAGGTGGCCCGAG 1052
901 GACCATGAGAGCTTCGCAAGACAGCTTTGAGCTTACGGGGGTGACTTCGCTCTTGGG 960
1053 GACCATGAGAGCTTCGCAAGACAGCTTTGAGCTTACGGGGGTGACTTCGCTCTTGGG 1112
961 AGGAGCTTCAGGCGCTGGCTGATCGAGATCAATTCAGAGCCGCCATGACACCGCTCCACG 1020
1113 AGGAGCTTCAGGCGCTGGCTGATCGAGATCAATTCAGAGCCGCCATGACACCGCTCCACG 1172
1021 CCGGTTCAGCGCCAGCTGTGTGACAGGTGACAGGAGACACCATCAAGGTGGCGCTGGAC 1080
1173 CCGGTTCAGCGCCAGCTGTGTGACAGGTGACAGGAGACACCATCAAGGTGGCGCTGGAC 1232
1081 CGAGCTGTGACATCGCAACTTCGAGCTCCTGTGTGAGGAGCGCGGTGTGTGAGCGGCC 1140
1233 CGAGCTGTGACATCGCAACTTCGAGCTCCTGTGTGAGGAGCGCGGTGTGTGAGCGGCC 1292
1141 CCATTGAGCGGTTCGCAACTTCGAGCTCCTGTGTGAGGAGCGCGGTGTGTGAGGAGCG 1200

1293 CCATTGAGCGGTTCGCAACTTCGAGCTCCTGCGTGGCGGCGTCAAGTGTGAGGAGCCAGGAGCGAG 1352
1201 GTCTCTCCCGCTGTGCAAACTCAAGGGCTTCGGCTTCGCTGTGTGACAGCGGAGCGCTGGAAG 1260
1353 GTCTCTCCCGCTGTGCAAACTCAAGGGCTTCGGCTTCGCTGTGTGACAGCGGAGCGCTGGAAG 1412
1261 GCACGGGGGCGCTTCGGCCATGCCAGACCTTGCCAGGGACCCCATCACCAGCTCTCCAG 1320
1413 GCACGGGGGCGCTTCGGCCATGCCAGACCTTGCCAGGGACCCCATCACCAGCTCTCCAG 1472
1321 CGGAGCTTGGGACTGAAGGAAGAGAGGGGCTCCCTCTGGCTTGTGGCACCCCTTAAAGG 1380
1473 CGGAGCTTGGGACTGAAGGAAGAGAGGGGCTCCCTCTGGCTTGTGGCACCCCTTAAAGG 1532
1381 GGGGCAAGCGAGAGCGGTGGAGCGGACAGCCCAACCGGACCAAAAGTGTGTGGGAAGGTG 1440
1533 GGGGCAAGCGAGAGCGGTGGAGCGGACAGCCCAACCGGACCAAAAGTGTGTGGGAAGGTG 1592
1441 GAGCTCCCGGCGCTTCGGCCATGCCAGACCTTGCCAGGGACCCCATCACCAGCTCTCCCG 1500
1593 GAGCTCCCGGCGCTTCGGCCATGCCAGACCTTGCCAGGGACCCCATCACCAGCTCTCCCG 1652
1501 GTAGCCAGCGCCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGGCAACCCGCTGGAGCCT 1560
1653 GTAGCCAGCGCCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGGCAACCCGCTGGAGCCT 1712
1561 GTGCTCGGGGCGCTGAAGACAGAGAGGGCGCGCTGCTCGCGCCCGCGGAGGAAAGGT 1620
1713 GTGCTCGGGGCGCTGAAGACAGAGAGGGCGCGCTGCTCGCGCCCGCGGAGGAAAGGT 1772
1621 TCA 1623
1773 TCA 1775

RESULT 3
US-10-615-659-12
; Sequence 12, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-12

Query Match 99.8%; Score 1619.8; DB 19; Length 3554;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1621; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGAGCGACGAGA 60
DB 1695 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGAGCGACAGA 1754
QY 61 AGCAAGCCAGGAGCCAGAGGAGGAGGCGGAGAGCAGCGACCTGAGCAGCAGGCAAGAT 120
DB 1755 AGCAAGCCAGGAGCCAGAGGAGGAGGCGGAGAGCAGCGACCTGAGCAGCAGGCAAGAT 1814
QY 121 GCTGAAAAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 180
DB 1815 GCTGAAAAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 1874
QY 181 AAGGTGTCCAGGCTTACCTGGGGAGCTGGAGCATGAGGACATCGACAGTCTCAGCAGAT 240

Db 1875 AAGGTGTGCGAGCCCTACCTGGGGAGCTGGAGCATGAGGACATCGACACGTGAGCAGAT 1934
|
Qy 241 GCCGTGAGGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTCCCTC 300
|
Db 1935 GCCGTGAGGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTCCCTC 1994
|
Qy 301 GTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATTAATTTTCGAGTGCCAGGCTCTG 360
|
Db 1995 GTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATTAATTTTCGAGTGCCAGGCTCTG 2054
|
Qy 361 CTGAATAGAAATCACTGTGTGAACCTTCAGACGGAATTTGACGGGCTCCGGAACATCTGG 420
|
Db 2055 CTGAATAGAAATCACTGTGTGAACCTTCAGACGGAATTTGACGGGCTCCGGAACATCTGG 2114
|
Qy 421 ATTATAAAGCCCGCGCAAGTCCCGGGCCGAGACATAGTGTGCATGACCGTGTGGAG 480
|
Db 2115 ATTATAAAGCCCGCGCAAGTCCCGGGCCGAGACATAGTGTGCATGACCGTGTGGAG 2174
|
Qy 481 GAGATCCTGGAGTGGCAGCTGCAGACCAACCTCTTTCCAGGGACAAACAAAGTGGGTGTC 540
|
Db 2175 GAGATCCTGGAGTGGCAGCTGCAGACCAACCTCTTTCCAGGGACAAACAAAGTGGGTGTC 2234
|
Qy 541 CAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTGACATGACAGATGG 600
|
Db 2235 CAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTGACATGACAGATGG 2294
|
Qy 601 TTCTCTCGTCAGGACCTGGAACCCCTGACCATCTGTTCTCAAGGAGAGTTACTTGGG 660
|
Db 2295 TTCTCTCGTCAGGACCTGGAACCCCTGACCATCTGTTCTCAAGGAGAGTTACTTGGG 2354
|
Qy 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAC 720
|
Db 2355 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAC 2414
|
Qy 721 GCCGTCCAGAAGTACCTGAAGAAATGATGTGGCGCAGCCCCCTGTGTCGCCGACACAAAC 780
|
Db 2415 GCCGTCCAGAAGTACCTGAAGAAATGATGTGGCGCAGCCCCCTGTGTCGCCGACACAAAC 2474
|
Qy 781 ATGTGACACAGACACAGGTTCCAGGAGTACTGACGCGCAGGCGCGTGGCCCGTGTGG 840
|
Db 2475 ATGTGACACAGACACAGGTTCCAGGAGTACTGACGCGCAGGCGCGTGGCCCGTGTGG 2534
|
Qy 841 GGCAGCGTTCATCTACCCGCTCCATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCAC 900
|
Db 2535 GGCAGCGTTCATCTACCCGCTCCATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCAC 2594
|
Qy 901 GACCAGTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTCTTGGG 960
|
Db 2595 GACCAGTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTCTTGGG 2654
|
Qy 961 AGGGACTTCAGGCCCCCTGGCTGATCGAGATCAATTCAGCCCCCACCATGACCCGTCACG 1020
|
Db 2655 AGGGACTTCAGGCCCCCTGGCTGATCGAGATCAATTCAGCCCCCACCATGACCCGTCACG 2714
|
Qy 1021 CCGGTCAAGCCCCAGCTGTGTGCACAGGTGACAGGAGACACCATCAAGTGGCCGCTGGAC 1080
|
Db 2715 CCGGTCAAGCCCCAGCTGTGTGCACAGGTGACAGGAGACACCATCAAGTGGCCGCTGGAC 2774
|
Qy 1081 CCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGAGAGCAGCCGGTGTGAGCGGCC 1140
|
Db 2775 CCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGAGAGCAGCCGGTGTGAGCGGCC 2834
|
Qy 1141 CCAATTCAGGGGTCCGACCTCTGCTGGCGGGCTCAGTGTGAGGAGACCCAGGAGGAC 1200
|
Db 2835 CCAATTCAGGGGTCCGACCTCTGCTGGCGGGCTCAGTGTGAGGAGACCCAGGAGGAC 2894
|
Qy 1201 GTGCTGCCGCTGTGCAACCTCAAGGCTCGGCTCGCTTTTGGACGCGCAGCGCTGAAG 1260
|
Db 2895 GTGCTGCCGCTGTGCAACCTCAAGGCTCGGCTCGCTTTTGGACGCGCAGCGCTGAAG 2954
|
Qy 1261 GCAAGGGGCCCCCTGGGCAATGCGAGACCTTGGCCAGGGAACCCCAATCAGCAGCTTCAG 1320
|

Db 2955 GCACGGGGCCCCCTCGGCCATGCGACACCTGCCCCAGGAGACCCCCCATCACAGCTCTCCAG 3014
|
Qy 1321 CGGGACTTTGGGACTCAAGGAAGAGAGGGGCTCCCCCTTGGCTGGCACCCCTTAAGG 1380
|
Db 3015 CGGGACTTTGGGACTCAAGGAAGAGAGGGGCTCCCCCTTGGCTGGCACCCCTTAAGG 3074
|
Qy 1381 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGGAAGGTG 1440
|
Db 3075 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGGAAGGTG 3134
|
Qy 1441 GAGCTCCCGGCTGCCCCCTGTGCGCACGTGGACAGTCAAGGCCCCCAACACCCGTGTCCCC 1500
|
Db 3135 GAGCTCCCGGCTGCCCCCTGTGCGCACGTGGACAGTCAAGGCCCCCAACACCCGTGTCCCC 3194
|
Qy 1501 GTAGCCCGAGCCCGCAAAAGCTGGGATCCAAACACAGCTAAATGCGCACCCGCTGGAGCCT 1560
|
Db 3195 GTAGCCCGAGCCCGCAAAAGCTGGGATCCAAACACAGCTAAATGAGGACCCGCTGGAGCCT 3254
|
Qy 1561 GTGCTGCGGGGCTTGAAGACAGCAGAGGCGCGCTGCGTCCGCCGCCGAGGAAAAGGT 1620
|
Db 3255 GTGCTGCGGAGCTTGAAGACAGCAGAGGCGCGCTGCGTCCGCCGCCGAGGAAAAGGT 3314
|
Qy 1621 TCA 1623
|
Db 3315 TCA 3317
|
RESULT 4
US-10-635-977-12
; Sequence 12, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-12

Query Match 99.8%; Score 1619.8; DB 19; Length 3554;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1621; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCATCCAGCATCCTCAAGTGGTGTGTCAGCCACCAAGCTGACGAGCTGACGAGGACGACAGA 60
|
Db 1695 ATGGCATCCAGCATCCTCAAGTGGTGTGTCAGCCACCAAGCTGACGAGCTGACGAGGACGACAGA 1754
|
Qy 61 AGCAAGCCCGAGGACACGAGGGAGGAGCGCGGAGCAGCGACCTGAGCAGGACGGAAGAT 120
|
Db 1755 AGCAAGCCCGAGGACACGAGGGAGGAGCGCGGAGCAGCGACCTGAGCAGGACGGAAGAT 1814
|
Qy 121 GCTGAAAATGTCTAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGGACATCGCGTGC 180
|
Db 1815 GCTGAAAATGTCTAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGGACATCGCGTGC 1874
|
Qy 181 AAGTGTGCGAGGCTTACTCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 240
|
Db 1875 AAGTGTGCGAGGCTTACTCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 1934
|
Qy 241 GCCGTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACCCAGCAGTACTTCCCTC 300
|
Db 1935 GCCGTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACCCAGCAGTACTTCCCTC 1994
|

QY 301 GTTCATGGCGATGTTTCATCTCCATTTCAAGAAATTACTTTTCGAGTGCAGAGGCTCTG 360
DB 1995 GTTCATGGCGATGTTTCATCTCCATTTCAAGAAATTACTTTTCGAGTGCAGAGGCTCTG 2054
QY 361 CTGAATAGAAATCAGCTCTGTGAACCCCTCAGACGGACATTCACGGGCTCCCGAAACATCTGG 420
DB 2055 CTGAATAGAAATCAGCTCTGTGAACCCCTCAGACGGACATTCACGGGCTCCCGAAACATCTGG 2114
QY 421 ATTATAAGCCCGGGCCAAAGTCCCGGGCCGAGACATAGTGTGCATGGAACCGTGTGGAG 480
DB 2115 ATTATAAGCCCGGGCCAAAGTCCCGGGCCGAGACATAGTGTGCATGGAACCGTGTGGAG 2174
QY 481 GAGATCTGAGCTGGCAGCTGCAGACCACTCTTTTCCAGGACCAACAGTGGGTGGTC 540
DB 2175 GAGATCTGAGCTGGCAGCTGCAGACCACTCTTTTCCAGGACCAACAGTGGGTGGTC 2234
QY 541 CAGAAGTACATCGAGACCGCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG 600
DB 2235 CAGAAGTACATCGAGACCGCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG 2294
QY 601 TTCCTCGTACGAGCTGGAAACCCCTGACCATCTGTTTACAGGAGATTACTTTCGGG 660
DB 2295 TTCCTCGTACGAGCTGGAAACCCCTGACCATCTGTTTACAGGAGATTACTTTCGGG 2354
QY 661 TTCCTAACTCAGCGCTTCTCCCTGGACAAGCTGGAAGCGCCATTCACCTGTGTGCAACAAC 720
DB 2355 TTCCTAACTCAGCGCTTCTCCCTGGACAAGCTGGAAGCGCCATTCACCTGTGTGCAACAAC 2414
QY 721 GCGCTGCAGAAATGACCTGAAGAATGATGTGGCGCGAGCCCTGCTGCCCGCACACAAC 780
DB 2415 GCGCTGCAGAAATGACCTGAAGAATGATGTGGCGCGAGCCCTGCTGCCCGCACACAAC 2474
QY 781 ATGTGACACGACACAGGTTTCCAGGAGTACTCTGCAGCGCCAGGCGCGTGGCGCGTGTGG 840
DB 2475 ATGTGACACGACACAGGTTTCCAGGAGTACTCTGCAGCGCCAGGCGCGTGGCGCGTGTGG 2534
QY 841 GGCAGGCTCATCTACCGTTCATGAAGAAGGCCCATTCGCCACGCCATGAAGTGGCCGAC 900
DB 2535 GGCAGGCTCATCTACCGTTCATGAAGAAGGCCCATTCGCCACGCCATGAAGTGGCCGAC 2594
QY 901 GACCAGCTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG 960
DB 2595 GACCAGCTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG 2654
QY 961 AGGGACTTCAGGCGCTTGGCTGATCGAGATCAATTCAGCGCCCAACCATGCACCCGTCACG 1020
DB 2655 AGGGACTTCAGGCGCTTGGCTGATCGAGATCAATTCAGCGCCCAACCATGCACCCGTCACG 2714
QY 1021 CCGGTCACGGCCCAGCTGTGTGCACAGGTGCAGGAGCAACCATCAAGTGTGGCCGTGGAC 1080
DB 2715 CCGGTCACGGCCCAGCTGTGTGCACAGGTGCAGGAGCAACCATCAAGTGTGGCCGTGGAC 2774
QY 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGGAGCGCGGTGTTGAGCGGCC 1140
DB 2775 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGGAGCGCGGTGTTGAGCGGCC 2834
QY 1141 CCATTACGCGGCTCCGACCTCTGCTGGCGGGCTCAGTGTGAGGAGAGCCAGAGGCGAG 1200
DB 2835 CCATTACGCGGCTCCGACCTCTGCTGGCGGGCTCAGTGTGAGGAGAGCCAGAGGCGAG 2894
QY 1201 GTGCTGCCCTTCGCAACCTCAAGGCTTCGGCTTCGCTGTGTGACGCGACGCGCTGGAAG 1260
DB 2895 GTGCTGCCCTTCGCAACCTCAAGGCTTCGGCTTCGCTGTGTGACGCGACGCGCTGGAAG 2954
QY 1261 GCACGGGGCCCTTCGGCCATGCGACACCTTCGCCAGGAGCCCCCATCACAGCTCTCCAG 1320
DB 2955 GCACGGGGCCCTTCGGCCATGCGACACCTTCGCCAGGAGCCCCCATCACAGCTCTCCAG 3014
QY 1321 CCGGACTTGGGACTGAAGGAGAGAGGGGCTCCCTTCGCTTGTGGCACCCCTTAAGG 1380
DB 3015 CCGGACTTGGGACTGAAGGAGAGAGGGGCTCCCTTCGCTTGTGGCACCCCTTAAGG 3074

QY 1381 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGGAAGGTG 1440
DB 3075 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGGAAGGTG 3134
QY 1441 GAGCTCCCGGCTTCGCCCTGTGCGCACAGTGGAGCAGTCAAGCCCAAAACACCGGTGTCCCC 1500
DB 3135 GAGCTCCCGGCTTCGCCCTGTGCGCACAGTGGAGCAGTCAAGCCCAAAACACCGGTGTCCCC 3194
QY 1501 GTAGCCAGCCCGCCAAAAGCTGGGATCCAAACCAAGCTAAATCGGCACCCGCTGGAGCCT 1560
DB 3195 GTAGCCAGCCCGCCAAAAGCTGGGATCCAAACCAAGCTAAATGAGCACCCGCTGGAGCCT 3254
QY 1561 GTGCTCGGGGCTGAAGACAGCAGAGGGCGGCTCGCTCCGCCCGCCGAGGAAAAGGT 1620
DB 3255 GTGCTCGGGGCTGAAGACAGCAGAGGGCGGCTCGCTCCGCCCGCCGAGGAAAAGGT 3314
QY 1621 TCA 1623
DB 3315 TCA 3317
RESULT 5
US-10-615-659-10
; Sequence 10, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-10
Query Match 92.8%; Score 1505.8; DB 19; Length 1859;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 2; Indels 47; Gaps 2;
QY 1 ATGGCATCCAGCATCTCTCAAGTGGGTGGTTCAGCCACAGAGCTGCAGAGGAGCAGCAGA 60
DB 47 ATGGCATCCAGCATCTCTCAAGTGGGTGGTTCAGCCACAGAGCTGCAGAGGAGCAGCAGA 106
QY 61 AGCAAGCCAGGAGCAGAGGAGGAGGCGCGGAGCAGCGACTGAGCAGAGGCAAGAT 120
DB 107 AGCAAGCCAGGAGCAGAGGAGGAGGCGCGGAGCAGCGACTGAGCAGAGGCAAGAT 166
QY 121 GCTGAAAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGCGTGC 180
DB 167 GCTGAAAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGCGTGC 226
QY 181 AAGGTGTGCGAGCCCTTACCTGGGCGAGCTGGAGATGAGGACATCCACAGCTCAGCAGAT 240
DB 227 AAGGTGTGCGAGCCCTTACCTGGGCGAGCTGGAGATGAGGACATCCACAGCTCAGCAGAT 286
QY 241 GCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAGCTGCACCCAGCAGTACTACTCCCTC 300
DB 287 GCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAGCTGCACCCAGCAGTACTACTCCCTC 346
QY 301 GTTCATGGCGATGCTTTTCAATTCAGAAATTAATCTTTTCGAGTGCAGGCTCTG 360
DB 347 GTTCA-----TGCAGGCTCTG 363
QY 361 CTGAATAGAAATCAGCTCTGTGAACCCCTCAGACGGACATTTGACGGGCTCCGGAAACATCTGG 420
DB 364 CTGAATAGAAATCAGCTCTGTGAACCCCTCAGACGGACATTTGACGGGCTCCGGAAACATCTGG 423

```
QY 421 ATTATAAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGATGGACCGGTGTGGAG 480
|
|
|
Db 424 ATTATAAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGATGGACCGGTGTGGAG 483
|
|
|
QY 481 GAGATCCTGGAGTGGCAGTGGACACACCTCTTTCCAGGGACAAAGTGGGTGGTTC 540
|
|
|
Db 484 GAGATCCTGGAGTGGCAGTGGACACACCTCTTTCCAGGGACAAAGTGGGTGGTTC 543
|
|
|
QY 541 CAGAAGTACATCAGACGCGCGTGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
|
|
|
Db 544 CAGAAGTACATCAGACGCGCGTGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 603
|
|
|
QY 601 TTCTCTCGTCAAGGACTGGAACCCCTTGACCATCTGTTTCTACAAGGAGAGTTACTTGGGG 660
|
|
|
Db 604 TTCTCTCGTCAAGGACTGGAACCCCTTGACCATCTGTTTCTACAAGGAGAGTTACTTGGGG 663
|
|
|
QY 661 TTCTCAACTCAGGGCTTCTCCCTGGACAAGCTGGACAGGGCCATCCACCTGTGCAACAAC 720
|
|
|
Db 664 TTCTCAACTCAGGGCTTCTCCCTGGACAAGCTGGACAGGGCCATCCACCTGTGCAACAAC 723
|
|
|
QY 721 GCGCTCCAGAAGTACCTGAAGAATGATGTGGGCGCGAGCCCTGCTGCGCCGACACAAC 780
|
|
|
Db 724 GCGCTCCAGAAGTACCTGAAGAATGATGTGGGCGCGAGCCCTGCTGCGCCGACACAAC 783
|
|
|
QY 781 ATGTGGACCAGCACCAAGTTTCCAGGAGTACTGTGAGCGCCAGGGCGTGGCGCGGTGTGG 840
|
|
|
Db 784 ATGTGGACCAGCACCAAGTTTCCAGGAGTACTGTGAGCGCCAGGGCGTGGCGCGGTGTGG 843
|
|
|
QY 841 GGCAGGCTCATCTACCCGTCTCATGAAGAAGGCATCGCCCAAGCCATGAAGGTGGCCGAC 900
|
|
|
Db 844 GGCAGGCTCATCTACCCGTCTCATGAAGAAGGCATCGCCCAAGCCATGAAGGTGGCCGAC 903
|
|
|
QY 901 GACCAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACCTTCCTTTGGG 960
|
|
|
Db 904 GACCAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACCTTCCTTTGGG 963
|
|
|
QY 961 AGGGACTTTCAGGCGCTTGGCTGATTCGAGATCAATTCAGCGCCCAACCATGACCCGTCACG 1020
|
|
|
Db 964 AGGGACTTTCAGGCGCTTGGCTGATTCGAGATCAATTCAGCGCCCAACCATGACCCGTCACG 1023
|
|
|
QY 1021 CCGGTCACGGCCAGCTGTGTGACAGGTGCGAGGAGACACATCAAGGTGGCGGTGGAC 1080
|
|
|
Db 1024 CCGGTCACGGCCAGCTGTGTGACAGGTGCGAGGAGACACATCAAGGTGGCGGTGGAC 1083
|
|
|
QY 1081 CGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGGAGGACGCGGTGGTTCAGCGCGCC 1140
|
|
|
Db 1084 CGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGGAGGACGCGGTGGTTCAGCGCGCC 1139
|
|
|
QY 1141 CCATTTCAGCGGGTTCGACCTCTCGTGGCGGGCGTCAAGTGTGAGGAGAGCCAGGAGGCAG 1200
|
|
|
Db 1140 CCATTTCAGCGGGTTCGACCTCTCGTGGCGGGCGTCAAGTGTGAGGAGAGCCAGGAGGCAG 1199
|
|
|
QY 1201 GTGTGCTCCGCTGTGCAACCTCAAGGCTCGGCTCGCTGTGTGACGCGAGCGGTGAAG 1260
|
|
|
Db 1200 GTGTGCTCCGCTGTGCAACCTCAAGGCTCGGCTCGCTGTGTGACGCGAGCGGTGAAG 1259
|
|
|
QY 1261 GCACGGGGCCCTCGGCGATGCAGACCTGCGCAGGGACCCCAACACAGCTTCACAG 1320
|
|
|
Db 1260 GCACGGGGCCCTCGGCGATGCAGACCTGCGCAGGGACCCCAACACAGCTTCACAG 1319
|
|
|
QY 1321 CGGGACTTGGGACTGAAGGAAGAGGGGCTCCCCCTTGGCTTGTGTGGACCCCTTAAGG 1380
|
|
|
Db 1320 CGGGACTTGGGACTGAAGGAAGAGGGGCTCCCCCTTGGCTTGTGTGGACCCCTTAAGG 1379
|
|
|
QY 1381 GGGGCAAGCCGAGAGCGGTGGAGCGGACAGCCCAACCCGACCAAGCTCTGGGAGGTG 1440
|
|
|
Db 1380 GGGGCAAGCCGAGAGCGGTGGAGCGGACAGCCCAACCCGACCAAGCTCTGGGAGGTG 1439
|
|
|
QY 1441 GAGCTCCGGGCTGCGCCCTGTGCGCACGTGGACAGTCAAGGCGCCCAACACCGGTGTCGCC 1500
|
|
|
Db 1440 GAGCTCCGGGCTGCGCCCTGTGCGCACGTGGACAGTCAAGGCGCCCAACACCGGTGTCGCC 1499
|
|
|
```

```
QY 1501 GTAGCCCGAGCCCGCCAAAGCTGGGATCCAAACCAAGCTAAATGCGCACCCCGCTGGAGCCT 1560
|
|
|
Db 1500 GTAGCCCGAGCCCGCCAAAGCTGGGATCCAAACCAAGCTAAATGAGCACCCCGCTGGAGCCT 1559
|
|
|
QY 1561 GTGCTGCGGGGCTGAAGACAGCAGAGGGCGGCTGCGTCCGCGCCCGGAGGAAAGCT 1620
|
|
|
Db 1560 GTGCTGCGGAGCCTGAAGACAGCAGAGGGCGGCTGCGTCCGCGCCCGGAGGAAAGGT 1619
|
|
|
QY 1621 TCA 1623
|
|
|
Db 1620 TCA 1622
|
|
|
RESULT 6
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10
Query Match 92.8%; Score 1505.8; DB 19; Length 1859;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 2; Indels 47; Gaps 2;
QY 1 ATGGCATCCAGCATCCTCAAGTGGGTGCTCAGCCACCAGAGCTGCAGAGGAGCAGCAGA 60
|
|
|
Db 47 ATGGCATCCAGCATCCTCAAGTGGGTGCTCAGCCACCAGAGCTGCAGAGGAGCAGCAGA 106
|
|
|
QY 61 AGCAAGCCCAAGGACACAGAGGAGGAGCGCGGAGCAGCGACTTGAGCAGAGCAAGAT 120
|
|
|
Db 107 AGCAAGCCCAAGGACACAGAGGAGGAGCGCGGAGCAGCGACTTGAGCAGAGCAAGAT 166
|
|
|
QY 121 GCTGAAAATGCTGAGGCAAAAGCTCAGGGGCTTCCCGGGGCGAGCTTGTGGACATCGCGTGC 180
|
|
|
Db 167 GCTGAAAATGCTGAGGCAAAAGCTCAGGGGCTTCCCGGGGCGAGCTTGTGGACATCGCGTGC 226
|
|
|
QY 181 AAGGTGTGCCAGGCTTACTCTGAGGCGAGCTGGAGCATGAGGACATCGACATCGACAGAT 240
|
|
|
Db 227 AAGGTGTGCCAGGCTTACTCTGAGGCGAGCTGGAGCATGAGGACATCGACATCGACAGAT 286
|
|
|
QY 241 GCGGTGGAGGACCTCACTGAGGCCCGAGTGGGAGGACCTTGACCCAGCAGTACTACTCCCTTC 300
|
|
|
Db 287 GCGGTGGAGGACCTCACTGAGGCCCGAGTGGGAGGACCTTGACCCAGCAGTACTACTCCCTTC 346
|
|
|
QY 301 GTTCATGCGCATGTCTTTCATCTCCAAATTCAGAAAATTACTTTTCGAGTGCAGGCTCTG 360
|
|
|
Db 347 GTTCA-----TGCCAGGCTCTG 363
|
|
|
QY 361 CTGAATAGATCACTGTGTGAACCTCAGACGGACATGACGGCTCCGGACATCTGG 420
|
|
|
Db 364 CTGAATAGATCACTGTGTGAACCTCAGACGGACATGACGGGCTCCGGACATCTGG 423
|
|
|
QY 421 ATTATAAAGCCCGCGCCCAAGTTCGCGGGCCGAGACATAGTGTGATGGACCGGTGTGGAG 480
|
|
|
Db 424 ATTATAAAGCCCGCGCCCAAGTTCGCGGGCCGAGACATAGTGTGATGGACCGGTGTGGAG 483
|
|
|
QY 481 GAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTTTCCAGGGACAAAGTGGGTGGTTC 540
|
|
|
```

Db	484		GAGATCCTGGAGCTGGCAGCTGCAGACCA	CCCTCTTTCAGGGGCAACAAGTGGGTGGTC	543
Qy	541	CAGAAGTAGCATCTGAGAGCCCGCTGCTCAT	CTGTGACACCAAGTTCGACATCAGACAGTGG	600	
Db	544	CAGAAGTACATCTGAGAGCGCGCTGCTCAT	CTGTGTGACACCAAGTTCGACATCAGACAGTGG	603	
Qy	601	TTCTCTGTCACGGACTGGAACCCCGCTGAC	CACTCTGGTTCTTCAAGAGAGATTACTTGGGG	660	
Db	604	TTCTCTGTCACGGACTGGAACCCCGCTGAC	CACTCTGGTTCTTCAAGAGAGAGTTACTTTCGGG	663	
Qy	661	TTCTCTCAACTCAGCGCTTCTCCCTGCGACA	AGCTGACAGCGGCATCCACCTGTGCAACAAC	720	
Db	664	TTCTCTCAACTCAGCGCTTCTCCCTGCGACA	AGCTGACAGCGGCATCCACCTGTGCAACAAC	723	
Qy	721	GCCTCTCAGAAGTACCTCGAAGATGATGT	TGGGCGCGACGCCCTGCTGCCGCAACAAC	780	
Db	724	GCCTCTCAGAAGTACCTCGAAGATGATGT	TGGGCGCGACGCCCTGCTGCCGCAACAAC	783	
Qy	781	ATGTGGACACGACACAGGTTCCAGGAGTGA	CTGTGAGCGCGCAGGCGCGTGGCGCGTGTGG	840	
Db	784	ATGTGGACACGACACAGGTTCCAGGAGTGA	CTGTGAGCGCGCGAGGCGCGTGGCGCGTGTGG	843	
Qy	841	GGCAGCGTCTATCTACCCGTCATGAAGAG	CCCATCGCCACGCCCATGAAGTGGGCCCAG	900	
Db	844	GGCAGCGTCTATCTACCCGTCATGAAGAG	CCCATCGCCACGCCCATGAAGTGGGCCCAG	903	
Qy	901	GACCACGTGGAGCCTTCGCAAGAACAGCT	TTTGAGCTCTACGGGGCTGACTTCGCTCTTGGG	960	
Db	904	GACCACGTGGAGCCTTCGCAAGAACAGCT	TTTGAGCTCTACGGGGCTGACTTCGCTCTTGGG	963	
Qy	961	AGGGACTTCAGGGCCTGGCTGATTCGAGAT	CAATTCAGCCGCCACCATGCAACCGTCCACG	1020	
Db	964	AGGGACTTCAGGGCCTGGCTGATTCGAGAT	CAATTCAGCCGCCACCATGCAACCGTCCACG	1023	
Qy	1021	CCGGTCAGCGCCGAGCTGTGTGACACAGG	TGCGAGGAGACACCATCAAGTGGCGCGTGAC	1080	
Db	1024	CCGGTCAGCGCCGAGCTGTGTGACACAGG	TGCGAGGAGACACCATCAAGTGGCGCGTGAC	1083	
Qy	1081	CGCAGCTGTGACATCTGGCAACTTCGAGCT	CTCTGTGGAGCGACGCGGTGGTTGAGCGCGCC	1140	
Db	1084	CGCAGCTGTGACATCTGGCAACTTCGAGCT	CTCTGTGGAGCGACGCGGTGGTTGAGCGCGCC	1139	
Qy	1141	CCATTCAGCGGGTCCGACCTCTGCGTGG	CGGGCGTCAGTGTGAGGAGAGCCAGGAGCAG	1200	
Db	1140	CCATTCAGCGGGTCCGACCTCTGCGTGG	CGGGCGTCAGTGTGAGGAGAGCCAGGAGCAG	1199	
Qy	1201	GTGCTGCCCGTCTGCAACCTCAAGGCC	CTCGGCTCGGCTGTGTGACGCGCAGCGCTGAAG	1260	
Db	1200	GTGCTGCCCGTCTGCAACCTCAAGGCC	CTCGGCTCGGCTGTGTGACGCGCAGCGCTGAAG	1259	
Qy	1261	GCACGGGGCCCTCGGCGATGCCAGAC	CCCTGCCCCAGGGACCCCCATCAACAGCTCTCCAG	1320	
Db	1260	GCACGGGGCCCTCGGCGATGCCAGAC	CCCTGCCCCAGGGACCCCCATCAACAGCTCTCCAG	1319	
Qy	1321	CGGCACTTGGGACTGAAGGAAGAGGGG	CTCCCTCGCTTGTGTGGCAACCCCTTAAGG	1380	
Db	1320	CGGCACTTGGGACTGAAGGAAGAGGGG	CTCCCTCGCTTGTGTGGCAACCCCTTAAGG	1379	
Qy	1381	GGGGCAGCCGAGCGGTGGAGCCGCA	CAGCCCAACCGCACAAGTGTCTGGGAAGGTG	1440	
Db	1380	GGGGCAGCCGAGCGGTGGAGCCGCA	CAGCCCAACCGCACAAGTGTCTGGGAAGGTG	1439	
Qy	1441	GAGCTCCGGCGCTGCCCTCTGTGCC	ACGCTGGAAGTCAAGACCGGGTGTCCCC	1500	
Db	1440	GAGCTCCGGCGCTGCCCTCTGTGCC	ACGCTGGAAGTCAAGACCGGGTGTCCCC	1499	
Qy	1501	GTAGCCCAGCCCGCAAAAGCTGGGAT	TCCAAACAGACTAAATGCACACCGCTGGAGCCT	1560	
Db	1500	GTAGCCCAGCCCGCAAAAGCTGGGAT	TCCAAACAGACTAAATGCACACCGCTGGAGCCT	1559	
Qy	1561	GTGCTCGGGGCTGTGAAGACAGAC	AGGGCGCGCTGCGTCCGCGCCCGGAGGAAAGGT	1620	

Db	1560	GTGCTCGGAGCCTGAAGACAGCAGAGGGCGCGCTGGCTCGCGCCGCGAGGAAAGGT	161
Qy	1621	TCA 1623	
Db	1620	TCA 1622	
RESULT 7			
US-10-615-659-11			
; Sequence 11, Application US/10615659			
; Publication No. US20040157234A1			
; GENERAL INFORMATION:			
; APPLICANT: Bristol-Myers Squibb Company			
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN			
; FILE REFERENCE: D0283 NP			
; CURRENT APPLICATION NUMBER: US/10/615,659			
; CURRENT FILING DATE: 2003-07-09			
; PRIOR APPLICATION NUMBER: U.S. 60/394,725			
; PRIOR FILING DATE: 2002-07-09			
; NUMBER OF SEQ ID NOS: 102			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 11			
; LENGTH: 3465			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-615-659-11			
Query Match 90.9%; Score 1475.8; DB 19; Length 3465;			
Best Local Similarity 95.7%; Pred. No. 0;			
Matches 1554; Conservative 0; Mismatches 2; Indels 67; Gaps 1;			
Qy	1	ATGGCATCCAGCATCTCTCAAGTGGTGGTCTCAGCCACCAGAGCTGCAGCAGGACGACGAGA 60	
Db	1678	ATGGCATCCAGCATCTCTCAAGTGGTGGTCTCAGCCACCAGAGCTGCAGCAGGACGACGAGA 1737	
Qy	61	AGCAAGCCCAAGGACACAGAGGGAGAGGCGCGGAGAGCGACCTGTAGCAGCAGGCAAGAT 120	
Db	1738	AGCAAGCCCAAGGACACAGAGGGAGAGGCGCGGAGAGCGACCTGTAGCAGCAGGCAAGAT 1797	
Qy	121	GCTGAAATCTCNGCMAAGCTCAGGGCGCTCCCGGGCAGCTTGTGGACATCGCGTGC 180	
Db	1798	GCTGAAATCTGAGGCAAGCTCAGGGCGCTCCCGGGCAGCTTGTGGACATCGCGTGC 1857	
Qy	181	AAGGTGTGCAGGCGCTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 240	
Db	1858	AAGGTGTGCAGGCGCTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 1917	
Qy	241	GCGGTGAGGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 300	
Db	1918	GCGGTGAGGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977	
Qy	301	GTTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGTCCGAGGCTCTG 360	
Db	1978	GTTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGTCCGAGGCTCTG 2037	
Qy	361	CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGGACATTTGACGGGCTCCGGAAACATCTGG 420	
Db	2038	CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGGACATTTGACGGGCTCCGGAAACATCTGG 2097	
Qy	421	ATTATAAGCCCGCGCCAAAGTCCCGGGGCCGAGACATAGTGTGCAATGGACCGTGTGGAG 480	
Db	2098	ATTATAAGCCCGCGCCAAAGTCCCGGGGCCGAGACATAGTGTGCAATGGACCGTGTGGAG 2129	
Qy	481	GAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAAACAGTGGGTGGTC 540	
Db	2130	-----AGGACAAACAGTGGGTGGTC 2150	
Qy	541	CAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTTCGACATTCAGACAGTGG 600	
Db	2151	CAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTTCGACATTCAGACAGTGG 2210	
Qy	601	TTCTCGTCAGGACTGGAAACCCCTCTGACCATTCTGGTTCTTACAAGGAGAGTTACTTGC GG 660	

|||||
2211 TTCTCTCGTCAGGACTGGAAACCCCTTGACCATCTGGTTCTACAAGAGAGATTACTTGGCG 2270
|||||
661 TTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGACAGCGCCATCCACCTGTGTCAACAAC 720
|||||
2271 TTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGACAGCGCCATCCACCTGTGTCAACAAC 2330
|||||
721 GCGCTCCAGAAGTACCTGAAGAAATGATGTGGGCGCAGCGCCCTCTGTGCGCGCACACAAC 780
|||||
2331 GCGCTCCAGAAGTACCTGAAGAAATGATGTGGGCGCAGCGCCCTCTGTGCGCGCACACAAC 2390
|||||
781 ATGTGACACAGACAGGTTCCAGGATGACCTGACAGCGCCAGCGCGCTGTGG 840
|||||
2391 ATGTGACACAGACAGGTTCCAGGATGACCTGACAGCGCCAGCGCGCTGTGG 2450
|||||
841 GGCAGCGTCTATCTACCCGTCCTATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCCCAG 900
|||||
2451 GGCAGCGTCTATCTACCCGTCCTATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCCCAG 2510
|||||
901 GACCAGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGCTGACTTCGTCTTGG 960
|||||
2511 GACCAGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGCTGACTTCGTCTTGG 2570
|||||
961 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACACATGACCCGTCACG 1020
|||||
2571 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACACATGACCCGTCACG 2630
|||||
1021 CCGGTCACGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGCGCGTGAAC 1080
|||||
2631 CCGGTCACGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGCGCGTGAAC 2690
|||||
1081 CGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1140
|||||
2691 CGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGAGGAGCAGCGGTGTTGAGCGGCC 2750
|||||
1141 CCAATTGAGGGGTCGACCTCTGCGTGGCGGGTCAAGTGTGAGGAGGACCGAGAGGAG 1200
|||||
2751 CCAATTGAGGGGTCGACCTCTGCGTGGCGGGCTCAAGTGTGAGGAGGACCGAGAGGAG 2810
|||||
1201 GTGCTCCGCTGTCAACCTCAAGGCTCGGCTCGCTGTGAGCGCGAGCGGCTGAAG 1260
|||||
2811 GTGCTCCGCTGTCAACCTCAAGGCTCGGCTCGCTGTGAGCGCGAGCGGCTGAAG 2870
|||||
1261 GCACGGGGCCCTCGGCCATGCCAGACCTTGCCAGGGACCCCAATCACCAGCTCTCCAG 1320
|||||
2871 GCACGGGGCCCTCGGCCATGCCAGACCTTGCCAGGGACCCCAATCACCAGCTCTCCAG 2930
|||||
1321 CGGGAATTGGGACTGAAGGAAGAGGGGCTCCCTCGGCTTGTGCGACCCCTTAAGG 1380
|||||
2931 CGGGAATTGGGACTGAAGGAAGAGGGGCTCCCTCGGCTTGTGCGACCCCTTAAGG 2990
|||||
1381 GGGGACGCGAGAGCGGTGGAGCGGCACAGCCACCCGACCAAGCTGCTGGGAGGTG 1440
|||||
2991 GGGGACGCGAGAGCGGTGGAGCGGCACAGCCACCCGACCAAGCTGCTGGGAGGTG 3050
|||||
1441 GAGCTCCCGGCTGCGCCCTGTGCGCACGTGGAGCAGTCAAGCGCCCAAAACCCGCTGTC 1500
|||||
3051 GAGCTCCCGGCTGCGCCCTGTGCGCACGTGGAGCAGTCAAGCGCCCAAAACCCGCTGTC 3110
|||||
1501 GTAGCCACGCGCCCAAAAGCTGGGATCCAAACAGCTAAATGCGCACCCGCTGAGGCT 1560
|||||
3111 GTAGCCACGCGCCCAAAAGCTGGGATCCAAACAGCTAAATGAGCACCCGCTGAGGCT 3170
|||||
1561 GTGCTCGGGGCTGAAGACAGCAGAGGGCGGCTGCGTCCGCGCCGAGGAAAGT 1620
|||||
3171 GTGCTCGGGGCTGAAGACAGCAGAGGGCGGCTGCGTCCGCGCCGAGGAAAGT 3230
|||||
1621 TCA 1623
|||||
3231 TCA 3233
|||||

US-10-635-977-11
; Sequence 11, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-11

Query Match 90.9%; Score 1475.8; DB 19; Length 3465;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1554; Conservative 0; Mismatches 2; Indels 67; Gaps 1;

QY 1 ATGGCATCCAGCATCTCTCAAGTGGGTGGTGCAGCCACAGAGCTGCAGAGGAGCAGCAGA 60
DB 1678 ATGGCATCCAGCATCTCTCAAGTGGGTGGTGCAGCCACAGAGCTGCAGAGGAGCAGCAGA 1737
QY 61 AGCAAGCCAGGGAACAGAGGAGGAGCGCGGAGCAGCGACTGAGCAGCGGCAAGAT 120
DB 1738 AGCAAGCCAGGGAACAGAGGAGGAGCGCGGAGCAGCGACTGAGCAGCGGCAAGAT 1797
QY 121 GCTGAATGCTGAGGCAAGCTCAGGGGCTCCGGGCGAGCTTGTGACATCGCGTGC 180
DB 1798 GCTGAATGCTGAGGCAAGCTCAGGGGCTCCGGGCGAGCTTGTGACATCGCGTGC 1857
QY 181 AAGGTGTGCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATGACACATCAGCAGAT 240
DB 1858 AAGGTGTGCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATGACACATCAGCAGAT 1917
QY 241 GCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGACCTGACCCAGCAGTACTCCCTC 300
DB 1918 GCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGACCTGACCCAGCAGTACTCCCTC 1977
QY 301 GTTCATGGCGATGCTTTCATCTCCAATTCAGAAATTAATTTCCAGTCCGAGGCTCTG 360
DB 1978 GTTCATGGCGATGCTTTCATCTCCAATTCAGAAATTAATTTCCAGTCCGAGGCTCTG 2037
QY 361 CTGAATAGAAATCACTCTGTGAACCTCAGACGCGACATTTGACGGGCTCCGGAAACATCTGG 420
DB 2038 CTGAATAGAAATCACTCTGTGAACCTCAGACGCGACATTTGACGGGCTCCGGAAACATCTGG 2097
QY 421 ATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGATGAGCCCGTGTGGAG 480
DB 2098 ATTATAAGCCCGCGCCAAAGTCCCGGGCCG----- 2129
QY 481 GAGATCTGGAGCTGGCAGCTGCAGACACCCCTCTTTCAGGGACAAACAAGTGGGTGTC 540
DB 2130 ----- 2150
QY 541 CAGAAGTACATCAGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
DB 2151 CAGAAGTACATCAGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 2210
QY 601 TTCCTCGTCA CGGACTGGAAACCCCTGACCATCTGTTCTTA CAAGAGAGTTACTTGGCG 660
DB 2211 TTCCTCGTCA CGGACTGGAAACCCCTGACCATCTGTTCTTA CAAGAGAGTTACTTGGCG 2270
QY 661 TTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGCAACAAC 720
DB 2271 TTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGCAACAAC 2330


```
QY 841 GGACGCGTTCATACCCGCTCATGAAGAGGCGCATCGCCACGCCATGAAGTGGCCCCAG 900
Db |||
QY 901 GACACGTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG 960
Db |||
QY 961 AGGGAGCTTCAGGCGCTCGAGATCAATTCAGAGCCACCATGACCGTCCAGC 1020
Db |||
QY 1046 AGGAGCTTCAGGCGCTCGAGATCAATTCAGAGCCACCATGACCGTCCAGC 1105
QY 1021 CCGGTTCAGCGCCAGCTGTGTGACAGGTGCAGAGGACACCATCAAGTGGCGGTGGAC 1080
Db |||
QY 1106 CCGGTTCAGCGCCAGCTGTGTGACAGGTGCAGAGGACACCATCAAGTGGCGGTGGAC 1165
QY 1081 CGGAGCTGTGACATCGGCAACTTCGAGCTCTGTGAGGAGCGCGTGTGAGCGCCGCC 1140
Db |||
QY 1166 CGGAGCTGTGACATCGGCAACTTCGAGCTCTGTGAGGAGCGCGTGTGAGCGCCGCC 1225
QY 1141 CCATTTCAGCGGCTCGGCTTCGAGTTCGAGGCGCGTTCAGTGTGAGGAGCGCGGAGCAG 1200
Db |||
QY 1226 CCATTTCAGCGGCTCGGCTTCGAGTTCGAGGCGCGTTCAGTGTGAGGAGCGCGGAGCAG 1285
QY 1201 GTGCTGCCGCTCTGCAACCTCAAGGCTTCGCGCTCGCTGTGAGCGCGAGCGCTGAAG 1260
Db |||
QY 1286 GTGCTGCCGCTCTGCAACCTCAAGGCTTCGCGCTCGCTGTGAGCGCGAGCGCTGAAG 1345
QY 1261 GCACGGGCGCTTCGCGCTATGCGAGCCCTGCGCGAGGAGCCCGCATCACAGCTTCCAG 1320
Db |||
QY 1346 GCACGGGCGCTTCGCGCTATGCGAGCCCTGCGCGAGGAGCCCGCATCACAGCTTCCAG 1405
QY 1321 CGGGAGCTTCGGAGTGAAGAGAGAGGGGCTCCCGCTGCTGCGCACCCCTTAAGG 1380
Db |||
QY 1406 CGGGAGCTTCGGAGTGAAGAGAGAGGGGCTCCCGCTGCTGCGCACCCCTTAAGG 1465
QY 1381 GGGGAGCGCGAGCGGTGGAGCGCGACAGCCCGCACCAAGCTGCTGGAGAGGTG 1440
Db |||
QY 1466 GGGGAGCGCGAGCGGTGGAGCGCGACAGCCCGCACCAAGCTGCTGGAGAGGTG 1525
QY 1441 GAGCTCCGCGCTTCGCGCTATGCGAGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCG 1500
Db |||
QY 1526 GAGCTCCGCGCTTCGCGCTATGCGAGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCG 1585
QY 1501 GTAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
Db |||
QY 1586 GTAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1645
QY 1561 GTGCTCGGGGCGCTGAAGACAGCAGAGGCGCGCTGCGTCCGCGCGCGCGCGCGCGCGCG 1620
Db |||
QY 1646 GTGCTCGGGGCGCTGAAGACAGCAGAGGCGCGCTGCGTCCGCGCGCGCGCGCGCGCG 1705
QY 1621 TCA 1623
Db |||
QY 1706 TCA 1708
```

RESULT 10

```
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
```

```
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match      90.8%; Score 1474.2; DB 19; Length 1939;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1553; Conservative 0; Mismatches 3; Indels 67; Gaps 1;

QY 1 ATGGCATCCAGCATCTCTCAAGTGGGTGTCAGCCACAGAGCTGAGAGGAGGAGCAG 60
Db ATGGCATCCAGCATCTCTCAAGTGGGTGTCAGCCACAGAGCTGAGAGGAGGAGCAG 212
QY 61 ACNAGCCCGAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
Db AGCAAGCCCGAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 272
QY 121 GCTGAAATGCTGAGGCAAAAGCTCAGGGGCTTCCCGGGGCGAGCTTGTGACATCGCGTGC 180
Db GCTGAAATGCTGAGGCAAAAGCTCAGGGGCTTCCCGGGGCGAGCTTGTGACATCGCGTGC 332
QY 181 AAGGTGTGCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 240
Db AAGGTGTGCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 392
QY 241 GCGGTGGAGGACCTCACTGAGGCGCGAGTGGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db GCGGTGGAGGACCTCACTGAGGCGCGAGTGGGAGGAGGAGGAGGAGGAGGAGGAG 452
QY 301 GTTCTATGGGATGCTTTTCAATCTCCAATTCAAGAAATTAATTTTCGAGTGCCAGGCTCTG 360
Db GTTCTATGGGATGCTTTTCAATCTCCAATTCAAGAAATTAATTTTCGAGTGCCAGGCTCTG 512
QY 361 CTGAATAGAAATCACGCTGTGTGAAACCTCAGACGGACATTTGACGGGTTCGGAACATCTGG 420
Db CTGAATAGAAATCACGCTGTGTGAAACCTCAGACGGACATTTGACGGGTTCGGAACATCTGG 572
QY 421 ATTATTAAGCCCGCGCGCGCAAGTCCCGGGGCGAGACATAGTGTGCATGACCGTGTGGAG 480
Db ATTATTAAGCCCGCGCGCGCAAGTCCCGGGGCGCG----- 604
QY 481 GAGATCTCTGAGCTGCGAGCTGCAGACCAACCTCTTTCCAGGAGCAACAAAGTGGGTGGTC 540
Db GAGATCTCTGAGCTGCGAGCTGCAGACCAACCTCTTTCCAGGAGCAACAAAGTGGGTGGTC 625
QY 541 CAGAAAGTACATCGAGACGCGCTGTCTATCTGTGACACCAAGTTGACATCGACAGTGG 600
Db CAGAAAGTACATCGAGACGCGCTGTCTATCTGTGACACCAAGTTGACATCGACAGTGG 685
QY 601 TTCTCTCGTCACGAGCTGGAAACCCCTGACCATCTGGTTCTACAAGGAGAGTACTTTGCGG 660
Db TTCTCTCGTCACGAGCTGGAAACCCCTGACCATCTGGTTCTACAAGGAGAGTACTTTGCGG 745
QY 661 TTCTCAAATCTCAGCGCTTCTCCTTGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 720
Db TTCTCAAATCTCAGCGCTTCTCCTTGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 805
QY 721 GCCGTCCAGAAATGACTGAAGAAATGATGTGGCGCGAGCGCCCTGTGTCGCCGACACAAC 780
Db GCCGTCCAGAAATGACTGAAGAAATGATGTGGCGCGAGCGCCCTGTGTCGCCGACACAAC 865
QY 781 ATGTGGACCGACACAGGTTCCAGGATGATCTGACGCGCGAGGCGCGTGGCGCGGTGG 840
Db ATGTGGACCGACACAGGTTCCAGGATGATCTGACGCGCGAGGCGCGTGGCGCGGTGG 925
QY 841 GGCAGCGCTCATCTACCCGCTCCATGAAGAGGCGCATCGCCACGCCCATGAAGTGGCGCCAG 900
Db GGCAGCGCTCATCTACCCGCTCCATGAAGAGGCGCATCGCCACGCCCATGAAGTGGCGCCAG 985
QY 901 GACCACGTGGAGCGCTCGCAAGAACAGCTTTGAGCTCTACGGGCTGACTTCGTCCTTGGG 960
```

Db 986 GACCACTGGAGCCTCGACAGAAAGCTTTGAGCTCTACGGGGTGACTTCTGCTCTGGG 1045
Qy 961 AGGACTTCAGGCCCTCGCTGATCGAGATCAATTCAGGCCCAACATGCAACCGCTCCAGC 1020
Db 1046 AGGACTTCAGGCCCTCGCTGATCGAGATCAATTCAGGCCCAACATGCAACCGCTCCAGC 1105
Qy 1021 CCGGTACAGGCCCAAGCTGTGACAGGTGCGAGGAGCACCACATCAAGGTGGCGTGGAC 1080
Db 1106 CCGGTACAGGCCCAAGCTGTGACAGGTGCGAGGAGCACCACATCAAGGTGGCGTGGAC 1165
Qy 1081 CGCAGCTGTGACATCGGCAACTTCGAGCTCTCGTGGAGGAGCAGCGGTGAGCCGCC 1140
Db 1166 CGCAGCTGTGACATCGGCAACTTCGAGCTCTCGTGGAGGAGCAGCGGTGAGCCGCC 1225
Qy 1141 CCATTACAGCGGTTCGACCTTCGCTGGCGGGCGTCAAGTGTGAGGAGCAGCAGAGAGCAG 1200
Db 1226 CCATTACAGCGGTTCGACCTTCGCTGGCGGGCGTCAAGTGTGAGGAGCAGCAGAGAGCAG 1285
Qy 1201 GTGCTGCCCGCTTCGAACCTCAAGGCTTCGCGCTCGCTGTGGAGCGGCGAGCGCTGAAG 1260
Db 1286 GTGCTGCCCGCTTCGAACCTCAAGGCTTCGCGCTCGCTGTGGAGCGGCGAGCGCTGAAG 1345
Qy 1261 GCAGCGGCCCTTCGCGCTTCGAGACCTTCGCGCGAGGAGCAGCGCTTCGAGCTTCAG 1320
Db 1346 GCAGCGGCCCTTCGCGCTTCGAGACCTTCGCGCGAGGAGCAGCGCTTCGAGCTTCAG 1405
Qy 1321 CGGACTTTGGGACTGAAGAGAGAGAGGGGTTCCTCCCTGCGCTTGTGGCAACCTTAAAG 1380
Db 1406 CGGACTTTGGGACTGAAGAGAGAGAGGGGTTCCTCCCTGCGCTTGTGGCAACCTTAAAG 1465
Qy 1381 GGGGACGCGAGAGCGGTGGAGCGGCGACAGCGGAGCAGCGGAGGAGGAGGAGGAGG 1440
Db 1466 GGGGACGCGAGAGCGGTGGAGCGGCGACAGCGGAGCAGCGGAGGAGGAGGAGGAGG 1525
Qy 1441 GAGCTCCCGGCTTCGCGCTTCGCGAGCTGAGACAGTCAAGGAGGAGGAGGAGGAGGAGG 1500
Db 1526 GAGCTCCCGGCTTCGCGCTTCGCGAGCTGAGACAGTCAAGGAGGAGGAGGAGGAGGAGG 1585
Qy 1501 GTAGCCAGCGCCGCAAAAGCTGGATCCAAACAGCTAAATCGCACCCGCTGGAGCCT 1560
Db 1586 GTAGCCAGCGCCGCAAAAGCTGGATCCAAACAGCTAAATCGCACCCGCTGGAGCCT 1645
Qy 1561 GTGCTGGGGGCTTGAGACAGCAGAGGCGCGCTGCGTCCGCGCGCGGAGGAGGAGGAGG 1620
Db 1646 GTGCTGGGGGCTTGAGACAGCAGAGGCGCGCTGCGTCCGCGCGCGGAGGAGGAGGAGG 1705
Qy 1621 TCA 1623
Db 1706 TCA 1708

RESULT 11
US-10-615-659-3
; Sequence 3, Application US/10615659
; Publication No. US2004015734A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIORITY FILING DATE: 2003-07-09
; PRIORITY FILING DATE: 2003-07-09
; PRIORITY FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS

; LOCATION: (1)..(726)
US-10-615-659-3
Query Match 44.7%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.4e-189;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 397 ATTGACGGGCTCGGAAACATCTGGATTATAAAGCCCGGCGCAAGTCCCGGGGGCGGAGAC 456
Db 1 ATTGACGGGCTCGGAAACATCTGGATTATAAAGCCCGGCGCAAGTCCCGGGGGCGGAGAC 60
Qy 457 ATAGTGTGCATGACACCGGTGTGGAGGAGATCTCTGAGCTGGCAGCTGCAGACCAACCTCTT 516
Db 61 ATAGTGTGCATGACACCGGTGTGGAGGAGATCTCTGAGCTGGCAGCTGCAGACCAACCTCTT 120
Qy 517 TCCAGGAGCAACAAAGTGGGTGGTCCAGAAATACATCGAGACGCGCTGTCTCATCTGTGAC 576
Db 121 TCCAGGAGCAACAAAGTGGGTGGTCCAGAAATACATCGAGACGCGCTGTCTCATCTGTGAC 180
Qy 577 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAACGAGCTGGACCCCTGACCATCTGG 636
Db 181 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAACGAGCTGGACCCCTGACCATCTGG 240
Qy 637 TTCTACAGGAGAGTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 696
Db 241 TTCTACAGGAGAGTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 300
Qy 697 AGCGCCATCCACTGTGTGCAACAAACGCGCTCCAGAAATGATGTGGGCGCG 756
Db 301 AGCGCCATCCACTGTGTGCAACAAACGCGCTCCAGAAATGATGTGGGCGCG 360
Qy 757 AGCGCCCTTCGCTCCGCGACACAAACATGTGACACGAGCTTCCAGGAGTACCTGAC 816
Db 361 AGCGCCCTTCGCTCCGCGCGACACAAACATGTGACACGAGCTTCCAGGAGTACCTGAC 420
Qy 817 CGCAGCGGCGCTGCGCGCTGTGGGCGAGCTCATCTACCGCTCCATGAAGAAGGCCATC 876
Db 421 CGCAGCGGCGCTGCGCGCTGTGGGCGAGCTCATCTACCGCTCCATGAAGAAGGCCATC 480
Qy 877 GCCCAGCCCATGAAGGTGGCCAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 936
Db 481 GCCCAGCCCATGAAGGTGGCCAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 540
Qy 937 TAGCGGCTGACTTCTGCTTGGGAGGAGCTTTCAGGCGCTGGCTGATCGAGATCAATTC 996
Db 541 TAGCGGCTGACTTCTGCTTGGGAGGAGCTTTCAGGCGCTGGCTGATCGAGATCAATTC 600
Qy 997 AGCCCCACCATGACACCGCTCCAGCGCTCCAGCGCGAGCTGTGTGACAGGTCAGAG 1056
Db 601 AGCCCCACCATGACACCGCTCCAGCGCTCCAGCGCGAGCTGTGTGACAGGTCAGAG 660
Qy 1057 GACACCATGAAGTGGCGCTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGG 1116
Db 661 GACACCATGAAGTGGCGCTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGG 720
Qy 1117 AGGCAG 1122
Db 721 AGGCAG 726
RESULT 12
US-10-615-977-3
; Sequence 3, Application US/10635977
; Publication No. US2004017113A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIORITY FILING DATE: 2003-08-07
; PRIORITY FILING DATE: 2003-08-07
; PRIORITY FILING DATE: 2003-08-07
; PRIORITY FILING DATE: 2003-08-07
; PRIORITY FILING DATE: 2003-08-07

; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-635-977-3

Query Match 44.7%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.4e-189;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 ATTGACGGGCTCGGAACATCTGGATTATAAAGCCCGCGGCCAAGTCCCGGGGCCGAGAC 456
Db 1 ATTGACGGGCTCGGAACATCTGGATTATAAAGCCCGCGGCCAAGTCCCGGGGCCGAGAC 60
QY 457 ATAGTGTGATGACGACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACCAACCCCTTT 516
Db 61 ATAGTGTGATGACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACCAACCCCTTT 120
QY 517 TCCAGGGACAAAGAATGGGTGTCCAGAAATATATCGAGACGCGCTGTCTCATCTGTGAC 576
Db 121 TCCAGGGACAAAGAATGGGTGTCCAGAAATATATCGAGACGCGCTGTCTCATCTGTGAC 180
QY 577 ACCAAGTTGACATACAGACAGTGTTCCTCGTACAGGACTGGAAACCCCTGACCATCTGG 636
Db 181 ACCAAGTTGACATACAGACAGTGTTCCTCGTACAGGACTGGAAACCCCTGACCATCTGG 240
QY 637 TTCTACAAGGAGATTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 696
Db 241 TTCTACAAGGAGATTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 300
QY 697 AGCGCCATCCACTGTGCAACACGCGTCCAGAAGTACTTGAAGAATGATGTGGSCCG 756
Db 301 AGCGCCATCCACTGTGCAACACGCGTCCAGAAGTACTTGAAGAATGATGTGGSCCG 360
QY 757 AGCCCCCTGCTGCCGCACACAAATGTGGACAGACACAGGTTCCAGGAGTACCTGGAC 816
Db 361 AGCCCCCTGCTGCCGCACACAAATGTGGACAGACACAGGTTCCAGGAGTACCTGGAC 420
QY 817 CGCCAGGGCGTGGCGCGTGTGGGACGCTCATCTACCCGTCCATGAAGAGGCCATC 876
Db 421 CGCCAGGGCGTGGCGCGTGTGGGACGCTCATCTACCCGTCCATGAAGAGGCCATC 480
QY 877 GCCCAGCCCATGAAGTGGCCAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 936
Db 481 GCCCAGCCCATGAAGTGGCCAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 540
QY 937 TAGGGGCTGACTTCGTCTCTGGAGGAGATTTCAGGCCCTGCTGATCGAGATCAATTCC 996
Db 541 TAGGGGCTGACTTCGTCTCTGGAGGAGATTTCAGGCCCTGCTGATCGAGATCAATTCC 600
QY 997 AGCCCCACCATGACCCCGTCCAGCGGTCCAGGCCAGCTGTGTGCACAGGTGCAGGAG 1056
Db 601 AGCCCCACCATGACCCCGTCCAGCGGTCCAGGCCAGCTGTGTGCACAGGTGCAGGAG 660
QY 1057 GACACCATCAAGTGGCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 1116
Db 661 GACACCATCAAGTGGCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 720
QY 1117 AGGCAG 1122
Db 721 AGGCAG 726

RESULT 13
US-10-723-860-631/c
; Sequence 631, Application US/10723860

; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 631
; LENGTH: 101270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-631

Query Match 30.9%; Score 501; DB 20; Length 101270;
Best Local Similarity 100.0%; Pred. No. 2.9e-127;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1120 CAGCGGTGTGAGCGCCCATTCAGCGGTCCGACCTCTGCGTGGCGGCGTCAGT 1179
Db 43637 CAGCGGTGTGAGCGCCCATTCAGCGGTCCGACCTCTGCGTGGCGGCGTCAGT 43578
QY 1180 GTGAGGAGCCAGAGGAGGTGTCGCCGTCTGCACTCAAGGCTTCGCGCTCGCTG 1239
Db 43577 GTGAGGAGCCAGAGGAGGTGTCGCCGTCTGCACTCAAGGCTTCGCGCTCGCTG 43518
QY 1240 TTGGACGCGCAGCGCTGAAGCAAGCGGGCCCTCGGCCATGCCAGACCTTGCCAGGGA 1299
Db 43517 TTGGACGCGCAGCGCTGAAGCAAGCGGGCCCTCGGCCATGCCAGACCTTGCCAGGGA 43458
QY 1300 CCCCACATCAAGCTCTCCAGCGGACTTTGGGACTTGAAGAAAGAGGGGCTCCCCCTG 1359
Db 43457 CCCCACATCAAGCTCTCCAGCGGACTTTGGGACTTGAAGAAAGAGGGGCTCCCCCTG 43398
QY 1360 GCCTTGTGGCACCTTAAGGGGGCGCAGAGGGTGGAGCGCGCACAGCCACCCCGC 1419
Db 43397 GCCTTGTGGCACCTTAAGGGGGCGCAGAGGGTGGAGCGCGCACAGCCACCCCGC 43338
QY 1420 ACCAAAGCTGTGGGAAGGTGAGGCTCCCGGCTGCCCCCTGCGCCACGTGGACAGTCA 1479
Db 43337 ACCAAAGCTGTGGGAAGGTGAGGCTCCCGGCTGCCCCCTGCGCCACGTGGACAGTCA 43278
QY 1480 GCCCAAAACACCGGTGTCCCGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCAAGTA 1539
Db 43277 GCCCAAAACACCGGTGTCCCGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCAAGTA 43218
QY 1540 AATGGCACCCGCTGGAGCTGTGTCGGGGCCCTGAACACAGCAGAGGGCGGCTGCGT 1599
Db 43217 AATGGCACCCGCTGGAGCTGTGTCGGGGCCCTGAACACAGCAGAGGGCGGCTGCGT 43158
QY 1600 CCGCCCGCCCGAGGAAAAGGT 1620
Db 43157 CCGCCCGCCCGAGGAAAAGGT 43137

RESULT 14
US-10-275-595A-39
; Sequence 39, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda

APPLICANT: LAL, Presti
APPLICANT: YAO, Monique G.
APPLICANT: BANDMAN, Olga
APPLICANT: BURFORD, Neil
APPLICANT: BATRA, Sajeev
APPLICANT: KEARNEY, Liam
APPLICANT: POLICKY, Jennifer L.
TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
FILE REFERENCE: PF-0772 USN
CURRENT APPLICATION NUMBER: US/10/275,595A
PRIORITY FILING DATE: 2003-06-13
PRIORITY FILING DATE: 2000-05-05
PRIORITY FILING DATE: 2000-05-08
PRIORITY FILING DATE: 2000-05-07
PRIORITY FILING DATE: 2000-06-07
PRIORITY FILING DATE: 2000-06-21
PRIORITY FILING DATE: 2000-06-21
NUMBER OF SEQ ID NOS: 68
SOFTWARE: PERL Program
SEQ ID NO 39
LENGTH: 2380
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No: 2156553CB1
US-10-275-595A-39

Query Match 26.7%; Score 432.8; DB 18; Length 2380;
Best Local Similarity 59.1%; Pred. No. 9.1e-109;
Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;

QY 108 CAGCAGCAGATGCTGAAATCTGAGCAAAAGCTCAG-GGGCTCCCGGGCAGCTTG 166
DB 361 CAGGAGCAACAGCGCCCAAGAAACAGAGAAACCCAGGTGGTGTCCCAAGATTTG 420
QY 167 TGACATCGCGTCAAGGTGTGCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCG 226
DB 421 TGGATGAGCTCTGTGCTGCGAGGAGTACCTTAGCAACTTGGCCACATGACATCG 480
QY 227 ACACGTGACGAGATGCGGTGGAGGACCTTACAGGCGGAGTGGAGGACCTGACCTCAGC 286
DB 481 ACAAGGACCTGGAGGCGCCCTGTACTACCCCGAGGCGTGGTCCCTTCTCTCCAGC 540
QY 287 AGTACTACTCCCTGCTTCATGGCGATGCTTTCATCTCCATTCAGAAATTAATTTTCG 346
DB 541 GCTACTACCAAGTGGTCCAGAAAGGGGCGAGAACTCAGGCACTTCGACATCAGGTCAGC 600
QY 347 AGTGCCAGGCTGCTGCAATAGAAATCAGTCTGTGAACCTCAGACGACATTCAGCGGC 406
DB 601 GCTGTGAGGACATCTTCGACGACTGAGCGCTGGTGTACCCAGATAGACATGGAAGGG 660
QY 407 TCCGGAACTCTGGAATTAAGCCCGCCCAAGTCCCGGGCCGAGACATAGTGTGCA 466
DB 661 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGGACATCATGTGCA 720
QY 467 TGAACCGTGTGGAGGAGATCCTGGAGCTGCGACCTGACAGACCACTCTTTCAGGGACA 526
DB 721 TGGACCACTCTGGAGGAGATGCTGAAGCTGGTGAACCGGCAACCCCGTGGTGAAGAGCG 780
QY 527 ACAAGTGGGTGGTCCAGAAATGATCAGACGCGCTGCTCATCTGTGACCAACAAAGTTTCG 586
DB 781 GCAAGTGGGTGGTCCAGAAATGATGAGCGGCGCCCTCTCATCTTTGGACCAAGTTTG 840
QY 587 ACATCAGACAGTGGTTCCTCGTCAGGACCTGGAACCCCTGACCATCTGCTTCTACAAGG 646
DB 841 ACCTCAGACAGTGGTTCCTCGTGAACCTGACCTGGAACCCACTTACCGTGTGTTCTACCGG 900
QY 647 AGAGTTACTTGGCGTCTCAACTCAGGGCTTCCCTTGGCAAGAGCTGGACAGCGCCATCC 706

DB 901 ACAGCTATATCCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAACTTGGACAACTCAGTGC 960
QY 707 ACCTGTGCAACAAACGCGCTCCAGAAATGATGTGAAGAAATGATGTGGCCGCGAGCCCTCTGC 766
DB 961 ACCTGTGCAACAACTCCATCCAGAAACACCTGAGAACTCATGCCATCGGCATCCACTGC 1020
QY 767 TGCCTCCGACACAACTATGTGGACACGACACAGGTTCCAGAGTACCTTGCAGCGCCAGGGCC 826
DB 1021 TTCCGCGACACAACTATGTGTCTTAGCCAGAGGTTCCAGGCCCCACCTGTCAGAGATGGGT 1080
QY 827 GTGGCCCGTGTGGGCGAGCGTCTATCTACCCGTCCATGAAGAAGGCCATCGCCACGCCCA 886
DB 1081 CCCCAAATGCTTGGTCCACCATCATCTGCTGCGCATGAAGATGCTGTGATCCAGCAC 1140
QY 887 TGAAGGTGGCCCAAGACACGCTGAGCCCTGCAAGAACAGCTTTTGAGCTCTACGGGGCTG 946
DB 1141 TTCAGACCTCCCAAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTATGGGGCTG 1200
QY 947 ACTTCGTCTTGGAGGAGCTTTCAGGCGCTTGCATGATCGAGATCAATTCAGGCCCCACCA 1006
DB 1201 ACTTCGTGTTCGGGGAGGACTTTCAGGCGCTGCTGATTCAGATCAACGCGCAGCCCCACGA 1260
QY 1007 TGCACCCGTCACGCGGTCACGCGCCAGCTGTGTCACAGGTGCAAGGAGGACACCATCA 1066
DB 1261 TGGCACCTCCACAGCAGTCACTGCCCCGGCTCTGTGCTGGCGTGCAGCTGACACCTGC 1320
QY 1067 AGGTGGCC-----GTGGACCGCAGCTGTGACATCGCAACTTCGAGCTCCTGT 1114
DB 1321 GCGTGTCTATTGACCGGATGCTGGACCGCACTGTGACACAGGAGCTTTGAGCTCATCT 1380
QY 1115 GAGGAGCGCGGTGTGAGCGCGCCCAATTCAGCGGCTCCGACCTTCTGCTGGCGGGCG 1174
DB 1381 ATAAGCAGCGCTGCTGTGGAGTGCCTCAATATGTGGGCATTCGGGCTCTCTGTAGAGGCT 1440
QY 1175 TCAGTGTGAGGAGAGCAGGAGGAGTGTGCTGCTGCAACCTCAAGGCTTCGSCCT 1234
DB 1441 TCACATCAAGAAAGCCCATGGCGATGTGCTATCGCGGATGGGGGTCCGCCAGAGTCC 1500
QY 1235 CGCTGTGAGCGCGCAGCGCTGAAGGACGCGGCGCCCTCGGCCATGCCC-AGACCTTGC 1293
DB 1501 CTCTGCTGACCCAGCAGGCTCTGGGGAAGGCAAGGACTCGGGGATCCCTTACCCACAGT 1560
QY 1294 CAGGAGCCCCATCACCAGCTCTCAGCGGAGCTTGGAGTGAAGAGAGAGGGGCTC 1353
DB 1561 CAGCTTCTAGGAAGGCACTGGGGCCAGGAGCTGGGGCCACAGTGAAGAGCCAGTCTCCA 1620
QY 1354 CCCTGGCCTTGTGSCACCTTAAAGGGGCGAGCGGAGAGCGGTGGAGCGGCACAGCCC 1413
DB 1621 CTGCCACCACTTCAGCCCCCGGAAAGGGAAGAAAGGCAAGGCAAGGCGCCACAGCCC 1680
QY 1414 ACCCGCACCAAGCTGCTGGGAAGGTGGAGCTCCCGGC 1451
DB 1681 TGGTCTGCCCAATCTCTGGGAGTGGGATGCCCCAGC 1718

RESULT 15

US-10-029-386-24894/c
Sequence 24894, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: ABOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24894
LENGTH: 490

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022327.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.35
; OTHER INFORMATION: SWISSPROT HIT: P38160, EVALUE 5.00e-04
; OTHER INFORMATION: EST_HUMAN HIT: BB827753.1, EVALUE 3.00e-06
; OTHER INFORMATION: NT HIT: G114726145, EVALUE 8.00e-09
US-10-029-386-24894

Query Match	26.3%;	Score 427;	DB 16;	Length 490;
Best Local Similarity	100.0%;	Pred. No. 2.8e-107;		
Matches 427;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	696	CAGCGCCATCCACCTGTGCAACAAACGCCGTCAGAAAGTACCTGAAGAATGATGTGGCGCG	755	
DB	427	CAGCGCCATCCACCTGTGCAACAAACGCCGTCAGAAAGTACCTGAAGAATGATGTGGCGCG	368	
QY	756	CAGCCCCCTGCTGCCCGGACACAACATGTGGACCAAGTTCAGAGGTACCTGCA	815	
DB	367	CAGCCCCCTGCTGCCCGGACACAACATGTGGACCAAGTTCAGAGGTACCTGCA	308	
QY	816	GGCCACAGGGCCGTGGGCGCGTGTGGGGCAGCGTCATCTACCCGTCATGAAGAAGGCCAT	875	
DB	307	GGCCACAGGGCCGTGGGCGCGTGTGGGGCAGCGTCATCTACCCGTCATGAAGAAGGCCAT	248	
QY	876	CGCCCCACGCCATGAAGGTGGCCCGCAGGACCAACGCTGGAGCCCTCGCAAGAACAGCTTTGAGCT	935	
DB	247	CGCCCCACGCCATGAAGGTGGCCCGCAGGACCAACGCTGGAGCCCTCGCAAGAACAGCTTTGAGCT	188	
QY	936	CTACGGGGGTGACTTCGTCTCTTGGGAGGGACTTCAGGGCCCTGGCTGATCGAGATCAATTC	995	
DB	187	CTACGGGGGTGACTTCGTCTCTTGGGAGGGACTTCAGGGCCCTGGCTGATCGAGATCAATTC	128	
QY	996	CAGCCCCACCATGCACCCGTCACACCGCGTCACGCCCGCAGCTGTGTGCACAGGTGCAGGA	1055	
DB	127	CAGCCCCACCATGCACCCCGTCACACCGCGTCACGCCCGCAGCTGTGTGCACAGGTGCAGGA	68	
QY	1056	GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCACCTTCGAGCTCCTGTG	1115	
DB	67	GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCACCTTCGAGCTCCTGTG	8	
QY	1116	GAGGCAG 1122		
DB	7	GAGGCAG 1		

Search completed: September 24, 2005, 16:48:40
Job time : 1255.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 20:40:04 ; Search time 296.571 Seconds
(without alignments)
8954.609 Million cell updates/sec

Title: US-10-615-659-1_COPY_153_1775
Perfect score: 1623
Sequence: 1 atggcaccagcactctcaa.....cgcccgaggaaaggttca 1623

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	421.4	26.0	3001	4	US-09-620-312D-494
2	56	3.5	909	4	Sequence 494, App
3	56	3.5	1305	4	Sequence 5992, Ap
4	52.6	3.2	1272	4	Sequence 5951, Ap
5	52.6	3.2	1449	4	Sequence 2471, Ap
6	52.6	3.2	2790	4	Sequence 2200, Ap
7	50.4	3.1	1524	4	Sequence 2632, Ap
8	50.2	3.1	9039	4	Sequence 3, Appli
9	48.6	3.0	2277	1	Sequence 983, App
10	48.6	3.0	2277	1	Sequence 5, Appli
11	48.6	3.0	2277	1	Sequence 5, Appli
12	48.2	3.0	2004	1	Sequence 18, Appl
13	48.2	3.0	2004	2	Sequence 18, Appl
14	48.2	3.0	2004	2	Sequence 18, Appl
15	48.2	3.0	2004	2	Sequence 18, Appl
16	48.2	3.0	2004	2	Sequence 18, Appl
17	48.2	3.0	2004	2	Sequence 18, Appl
18	48.2	3.0	2004	3	Sequence 17, Appl
19	48.2	3.0	2655	1	Sequence 17, Appl
20	48.2	3.0	2655	2	Sequence 17, Appl
21	48.2	3.0	2655	2	Sequence 17, Appl
22	48.2	3.0	2655	2	Sequence 17, Appl
23	48.2	3.0	2655	2	Sequence 17, Appl
24	48.2	3.0	2655	2	Sequence 17, Appl
25	48.2	3.0	2655	3	Sequence 17, Appl
26	48	3.0	5314	4	Sequence 138, App
27	47.6	2.9	927	4	Sequence 9188, Ap

28	47.6	2.9	4884	4	US-09-252-991A-12126	Sequence 12126, A
29	47.6	2.9	4884	4	US-09-252-991A-12292	Sequence 12292, A
30	46.4	2.9	4600	4	US-09-702-705-1797	Sequence 1797, Ap
31	46.4	2.9	4600	4	US-09-736-457-1797	Sequence 1797, Ap
32	46.4	2.9	4600	4	US-09-671-325-1797	Sequence 1147, Ap
33	46.4	2.9	19455	4	US-09-902-540-1147	Sequence 1147, Ap
34	46	2.8	1218	4	US-09-902-540-5379	Sequence 5379, Ap
35	45	2.8	2847	4	US-09-902-540-7588	Sequence 7588, Ap
36	46	2.8	7704	4	US-09-902-540-743	Sequence 743, App
37	46	2.8	34552	4	US-09-902-540-1262	Sequence 1262, Ap
38	45.8	2.8	1515	4	US-09-902-540-9184	Sequence 9184, Ap
39	45.8	2.8	9556	4	US-09-902-540-982	Sequence 982, App
40	45.6	2.8	534	4	US-09-252-991A-7596	Sequence 7596, Ap
41	45.6	2.8	576	4	US-09-252-991A-7886	Sequence 7886, Ap
42	45.6	2.8	1047	4	US-09-252-991A-14678	Sequence 14678, A
43	45.6	2.8	1566	4	US-09-902-540-4578	Sequence 4578, Ap
44	45.6	2.8	1686	4	US-09-252-991A-14548	Sequence 14548, A
45	45.6	2.8	1695	4	US-09-252-991A-7816	Sequence 7816, Ap

ALIGNMENTS

RESULT 1
US-09-620-312D-494
; Sequence 494, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Felyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt.PL_genes Version 1.0
; SEQ ID NO 494
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1361)..(2419)
US-09-620-312D-494

Query Match 26.0%; Score 421.4; DB 4; Length 3001;
Best Local Similarity 53.8%; Pred. No. 1.6e-89;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 80 GGGAGGAGCGGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA 139
|||||
Db 1223 GGGATCGTGACAGGCCCTCCCTATGTGAGGAGACAGCAGCCCAAGAAACAGGAGAAA 1282
|||||

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5951

; LENGTH: 1305

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5951

Query Match 3.5%; Score 56; DB 4; Length 1305;

Best Local Similarity 44.4%; Pred. No. 0.0014;

Matches 224; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

Qy 780 CATGTGACACGACACAGGTTCACAGGAGTACTCTGACGCGCCAGCGCGTGGCGCGGTG 839

Db 546 CCGTATGAGCTTACCGCGCGCGCGGACAGAGAGGCGCGGTCCGGTGAAGGTGCG 605

Qy 840 GGGCAGCGGTTCATCTACCGCTTCATGAAGAAGGCCATCGCCCAACGCCATGAAGTGGCCCA 899

Db 606 CGTGGCGCTGACCGATATCTCTACCGGGCTGTACTCTCCACCGCGGTGCTCGCGCGCCT 665

Qy 900 GGACACGTGAGCGCTCGAAGAACAGCTTGTAGCTCTAGCGGCGTCACTTGTCTTGG 959

Db 666 CGCCACCGCGACGCTCAGCGGCATCGGCCAGCACATCGACATGGCCCTGTCTCGACGTGCA 725

Qy 960 GAGGACTTCAGGCGCTGCTGTCGAGATCAATTCCAGCGCCACCATGACCGTCCAC 1019

Db 726 GTGCGCTGCTGGCCAAACAGACCCCTCACTGACCAACCGCGGTGCGCGCGCGCG 785

Qy 1020 GCGCGTCACGCGCCAGCTGTGTGCACAGGTGTCAGGAGACACCATCAAGTGGCGGTGGA 1079

Db 786 CTTGGGCAACGACACCCGAACTCGTGGCTACCAAGACTTCCCAACCGCGAGCGGA 845

Qy 1080 CCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGAGGAGCGCGGTGTGAGCGCC 1139

Db 846 CATGATCTCAGCGTCGCGCAACGACAGCCAGTTCGCAAGTTCGCCAGCTGGCAGACCA 905

Qy 1140 CCAATTACGCGGTCGCACTCTCGTGGCGGGCGTCACTGAGGAGAGCCAGGAGCA 1199

Db 906 CCGGGAATGGCGGACAGACCGCGCTTCGCCAACCAAGCGCGGTGGCCAAACCGCGA 965

Qy 1200 GGTGCTGCGCGCTCTGCAACCTCAAGGCTCGCGCTGTGAGCGCGAGCGCTGAA 1259

Db 966 GTGCTGATCCCGTGTATCCGCGAGGCCACGGTCTGCACACCAACCGCGAGTCTCT 1025

Qy 1260 GGCACGGGCGCCCTCGGCCATGCC 1283

Db 1026 TTCCCTGGAGCGCGCGCGGTGCC 1049

RESULT 4

US-09-252-991A-2471

; Sequence 2471, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2471

; LENGTH: 1272

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-2471

Query Match

Best Local Similarity

3.2%; Score 52.6; DB 4; Length 1272;

48.5%; Pred. No. 0.0088;

Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

Qy 393 GGACATTGACGGGCTCGGAAACATCTGGATTATAAGCCCGCCAGTCCCGGGCGG 452

Db 756 GGAATGCTCGCGCAGCGCGGACGAGTTCGGCAGCCGCGAGGACGCGATCCT 815

Qy 453 AGACATAGTGTGATCGACCGTGTGGAGGAGATCTTGGAGCTGCAGCTGCAGACCAACC 512

Db 816 CGACGTACGATGTTGACACCGCGGACGCGCGCTGGAGTGGTGGAGTAGGCGAGGC 875

Qy 513 TCTTTCCAGGGACAACAAGTGGGTGTCAGAAAGTACATCGAGACGCGCTGTCTCATCTG 572

Db 876 GGTGACGCTGGAAGTTCGAGTTCGAGTTCGCCAGGACATCGAGCGGTGATCCTCGGCTT 935

Qy 573 TGACACCAAGTTGACATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCAT 632

Db 936 CATGATCAAGGACCGCTCGGCCAGCGATGTACGGGATCAATACCCACCGCTTGGACAA 995

Qy 633 CTGGTTCTACAAGGAGAGTTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 691

Db 996 GGCCTCACCGACTGAAGCCCGCGAGCGCATCACTACCGCTTCGCTTCGACATGC 1054

RESULT 5

US-09-252-991A-2200/c

; Sequence 2200, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2200

; LENGTH: 1449

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-2200

Query Match 3.2%; Score 52.6; DB 4; Length 1449;

Best Local Similarity 48.5%; Pred. No. 0.0092;

Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

Qy 393 GGACATTGACGGGCTCGGAAACATCTGGATTATAAGCCCGCCAGTCCCGGGCGG 452

Db 562 GGAATGCTCGCGCAGCGCGGACGAGTTCGGCAGCCGCGAGGACGCGATCCT 503

Qy 453 AGACATAGTGTGATCGACCGTGTGGAGGAGATCTTGGAGCTGCAGCTGCAGACCAACC 512

Db 502 CGACGTACGATGTTGACACCGCGGACGCGCGCTGGAGTGGTTCAGTAGGCGAGGC 443

Qy 513 TCTTTCCAGGGACAACAAGTGGGTGTCAGAAAGTACATCGAGACGCGCTGTCTCATCTG 572

Db 442 GGTGACGCTGGAAGTTCGAGTTCGAGTTCGCCAGGACATCGAGCGGTGATCCTCGGCTT 383

Qy 573 TGACACCAAGTTGACATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCAT 632

Db 382 CATGATCAAGGACCGCTCGGCCAGCGGATGTACGGGATCAATACCCACCGCTTGGACAA 323

Qy 633 CTGGTTCTACAAGGAGAGTTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 691

Db 322 GGCCTCACCGACTGAAGCCCGCGAGCGCATCACTACCGCTTCGCTTCGACATGC 264

RESULT 6

US-09-252-991A-2632

; Sequence 2632, Application US/09252991A

Qy	1149	CGGCTCGACCTCTGCTGCGGGCGCTAGTGTGAGGAGACCGAGGAGCAGGTGCTGCC	1208
Db	3048	AGCGTGTGTTGGTCTGCTGCTGCGCCCTCTCACTGACCCGCTCTCAAGGAAGACCCCGTGGG	3107
Qy	1209	CGTCTGAACC--TCAAGGCGCTCGGCTCTGCTGTGTGACGCGACGCGCTGAAGGCACCG	1266
Db	3108	AAGATGAGCCCATTTCCCCGCCCGCGGGATGAAGGACGCTCGCCTCCCGCGGAAGCC	3167
Qy	1267	GGCCCTCGGCATGCCAGACCTGCGCCAGGACCCCATCAACGACTCTCCAGCGGGAC	1326
Db	3168	GCTCCCCGGGCTGCCGCGGACCAACCGTGGACCCCTCAAAGGTGTGCTCTCCGCTGAAG	3227
Qy	1327	TTGGGACTGAAGGAAGAAGGGGCTCCCCCTGGCCTTGCTGGCACCCCTTAAGGGGGCA	1386
Db	3228	ATTGCCAGGCGCGCCCTCGTGTACCGGCTCGACATCGAGCGCACCGCGCGCGGCC	3287
Qy	1387	GCGGAGCGGTGGAGCGGCACAGCCACCGGCACCAAGCTGCTGGAGGTGGAGCTC	1446
Db	3288	GCCAGCACCGAGGAAGCCGGCGCCCCCAAGAGCGCCAAAGGGCAAGGCAAGGAAGCC	3347
Qy	1447	CCGCGCTGCCCTCTGTGCGCAACGTGGACAGTCAGGCCCCCAACACC	1491
Db	3348	GCCCCCTGCTGCGCGGGCCCCCGGGTCTGTGCGGGGCGCGCTACC	3392

RESIT.T 9

```

US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS: 10
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967

```

```

CLASSIFICATION: 330
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-676-967-5

```

	Query Match	3.0%	Score 48.6;	DB 1;	Length 2277;
	Best Local Similarity	49.8%;	Pred. No. 0.093;		
	Matches 123;	Conservative	0;	Mismatches 124;	Indels 0; Gaps 0;
Qy	681	CCTGGCAACCTGGACAGCGCCATCCACCTGTGTGCACAACGCCGTCTCAGAAGTACCTGAA	740		
Db	1437	CCAGAACATCTTCTGTAGCGCGACCCGCCCTGTGCTCTGCACAACTGCCAAGGCCGTGGA	1496		

Qy	741	G A N T G A T G T G G G C C G C A G C C C C T G C T G C C C G C A C A A C A T G T G A C C A G C A C C A G T T	800
Db	1497	C G A C A G C A G C T G C G C A A G C T G C T G T G A G C G C C A C C A G C G G C G A A G G C G T G C G C A T	1556
Qy	801	C C A G G A G T A C C T C A C G C C C A G G C C G T G G C G C G T G T G G G G C A G C G T C A T C T A C C G T C	860
Db	1557	C A A G G A G T C C G C G T G A T G C G C A C C T G A A G G C G C T G C A C G G C A C A T A G A G G C C A G A G	1616
Qy	861	C A T G A A G A A G G C C A T C G C C C A C G C C A T G A A G T G G C C C A G G A C C A C G T G G A G C C T C G C A A	920
Db	1617	C C T G G G C T A C G C C T T C G C C G A G T T C C A G G A G C A C A G A C A G C C C T G A A G G C C C T G C G C C T	1676
Qy	921	G A A C A G C	927
Db	1677	G A T C A A C	1683

RESULT 10

```

US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC896-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-676-974-5

```

Query Match 3.0%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.093;
Matches 123; Conservative 0; Mismatches 124; Indels 0

Qy	681	CCTGGACAAGCTGGACAGCGCCATCATCTGTGTGCAACAACGCCGTCCAGAAGTACTCTGAA	740
Db	1437	CCAGAACATCTTGTGAGCGCACCCCGCTGTGCTGTGCACAACTCTCCAAAGCCCGTGA	1496
Qy	741	GAATGATGTGGCGCGCAGCCCCCTGTGCCCCGACACAACATGTGACACAGACCAAGTTT	800
Db	1497	CGACAAGCAGCTGCGAAGTGTCTGTGAGCGCCACACAGCGCGAGAAGGGCGTGCGCAT	1556
Qy	801	CCAGGAGTACTTCAGCGGCCAGGGCCGTGGCCGCTGTGGGGCAGCGTCATCTACCCGTC	860
Db	1557	CAAGGAGTGC CGCGTGTGATGCGCACCTGAAGGCGCTGCACGGCAACATGAAGGGCCAGAG	1616

Qy 861 CATGAAGAGGCCATCGCCACGCGCATGAGGTGGCCAGGACCACCTGGAGCCTCGCAA 920
 |||||
Db 1517 CCTGGGTACGCCTTCGCCGAGTTCACAGGACACGAGCACGCCCTGAAGGCCCTGCGCCT 1676
 |||||
Qy 921 GAACAGC 927
 ||||
Db 1577 GATCAAC 1683

```

RESULT 11
US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-098-487-5

```

	Query Match	3.0%;	Score 48.6;	DB 2;	Length 2277;
	Best Local Similarity	49.8%;	Pred. No. 0.093;		
	Matches 123;	Conservative 0;	Mismatches 124;	Indels 0;	Gaps 0;
QY	681	CCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAACCGCTGCCAGAACTACCTGAA	740		
DB	1437	CGAAGACATCTTCTGTGAGCGCACCCGCCCTGTGCTGTGCACAACTGTGCCAAGGCCGTGGA	1496		
QY	741	GAATGATGTGGGCGCGACGCCCTGTCTGCCCGCACACAAATGTGGACACGACCACAGTTT	800		
DB	1497	CGACAAGCAGCTGCGCAAGCTGTCTGTGAGCGCCACCCAGCGGCGAGAAAGCGCGTGGCAT	1556		
QY	801	CCAGGAGTACTTCAGCGCCACAGGCGCGTGGCGCGTGTGGGGCAGCGTCACTACCGGTC	860		
DB	1557	CAAGAGTGCCGGTGATGTCGGACCTGAAGGGGTGCACGGCAACATGAAGGGCCAGAG	1616		
QY	861	CATGAAGAAGGCCATGCCCAACGCCATGAAGGTGGCCAGGACCAACGTGGAGCCTCGCAA	920		
DB	1617	CCTGGGCTACGCCCTTCCGCCAGTTTCCAGGAGCACGACAGCCCTGAAGGCCCTGGGCCT	1676		

```

RESULT 12
US-08-471-033-18
; Sequence 18, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB71
US-08-471-033-18

```

	Query Match	3.0%;	Score 48.2;	DB 1;	Length 2004;
	Best Local Similarity	43.7%;	Pred. No. 0.11;	273;	Indels 0
	Matches 212;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	507	CCACCCCTCTTTTCAGGACAA	CAAGTGGTGGTCCAGAGTAC	ATCGAGACGCCCGCTCT	566
Db	720	CCACCCCATCACCTCTGAAC	CAAGACAGGTGGCAACCTG	CTGAACAAACAGGCCCATGAT	779

QY 567 CATCTGTGACACCAAGTTTCAGATCATCAGACAGTGGTTCCTCGTCACGAGCTGGAAACCCCT 626
Db 780 GCTGGAGACCAACCAAGACCGGCGTCTCAAGATCAAGACACCCACCGGCAACATCGT 839
QY 627 GACCATCTGGTTTCAAGAGAGATTACTTTCGGTTTCTCAACTCAGCGCTTCTCCCTGGA 686
Db 840 GACCGCGCGAGTGGAAACCGGTGATCCAGCAGATCAAGGCCAAGACCGCCAGCATCAT 899
QY 687 CAAGCTGGACAGCGCCATCATCTGTGTCACCAAGCCGTCAGAAAGTACCTGAAGATGA 746
Db 900 CGTCGACGACGCGGAGCGCGTGGCCGAGAAAGCGGTGGCCGCAAGGACTACGAGAACCC 959
QY 747 TGTGGCGGACGCGCCCTGCTGCGCCGACACAAATGTGGACACGACCAAGTTTCCAGGA 806
Db 960 CGAGGACAAGACCCCAAGCTGACCTGAGAGAGCCCTGAAGCTGAGCTACCCCGACGA 1019
QY 807 GTACCTGCAGCGCCAGCGCGCTGTCGCGCGTGTGGGGCAGCGTCTATCTACCCGTCATGAA 866
Db 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACCAAGCCCATCTACGAGAGCAGCT 1079
QY 867 GAAGGCATCGCCCAACCCATGAAGTGGCCGAGGACCAAGTGGAGCGCTCGCAAGAACAG 926
Db 1080 GATGACCTATCTAGACGAGAACACCGCCAAAGAGGTGACCAAGCAGCTGAACGACACAC 1139
QY 927 CTTTGACCTTACGGGCTGACTTCTGTCCTTGGAGGAGCTTCAAGGCGCTGGCTGATCGA 986
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAGCTGACCCCAAGATGAACGT 1199
QY 987 GATCA 991
Db 1200 GACCA 1204

RESULT 13

US-08-471-044-18
; Sequence 18, Application US/08471044
; Patent No. 5840868

; GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868 Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,044

FILING DATE: 06-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/463,483

FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA: US 08/037,057
; APPLICATION NUMBER: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SQLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
; US-08-471-044-18

Query Match 3.0%; Score 48.2; DB 2; Length 2004;

Best Local Similarity 43.7%; Pred. No. 0.11;

Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 507 CCACCTCTTCCAGGACCAACAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCT 566
Db 720 CCACCCCATCACCTTGAACAAAGAGCGGTGGCAACCTCTGTAACAAACAGCCCATGAT 779
QY 567 CATCTGTGACACCAAGTTTCAGATCATCAGACAGTGGTTCCTCGTCACGAGCTGGAAACCCCT 626
Db 780 GCTGGAGACCAACCAAGACCGCGCTCTACAAGATCAAGGACACCCACCGCAACATCGT 839
QY 627 GACCATCTGGTTTCAAGAGAGATTACTTTCGGTTTCTCAACTCAGCGCTTCTCCCTGGA 686
Db 840 GACCGCGCGAGTGGAAACCGGTGATCCAGCAGATCAAGGCCAAGACCGCCAGCATCAT 899
QY 687 CAAGCTGGACAGCGCCATCATCTGTGTCACCAAGCCGTCAGAAAGTACCTGAAGATGA 746
Db 900 CGTCGACGACGCGGAGCGCGTGGCCGAGAAAGCGGTGGCCGCAAGGACTACGAGAACCC 959
QY 747 TGTGGCGGACGCGCCCTGCTGCGCCGACACAAATGTGGACACGACCAAGTTTCCAGGA 806
Db 960 CGAGGACAAGACCCCAAGCTGACCTGAGAGAGCCCTGAAGCTGAGCTACCCCGACGA 1019
QY 807 GTACCTGCAGCGCCAGCGCGCTGTCGCGCGTGTGGGGCAGCGTCTATCTACCCGTCATGAA 866
Db 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACCAAGCCCATCTACGAGAGCAGCT 1079
QY 867 GAAGGCATCGCCCAACCCATGAAGTGGCCGAGGACCAAGTGGAGCGCTCGCAAGAACAG 926
Db 1080 GATGACCTATCTAGACGAGAACACCGCCAAAGAGGTGACCAAGCAGCTGAACGACACCC 1139
QY 927 CTTTGACCTTACGGGCTGACTTCTGTCCTTGGAGGAGCTTCAAGGCGCTGGCTGATCGA 986
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAGCTGACCCCAAGATGAACGT 1199
QY 987 GATCA 991
Db 1200 GACCA 1204

RESULT 14

US-08-463-483A-18
; Sequence 18, Application US/08463483A
; Patent No. 5849870

; GENERAL INFORMATION:

APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-463-483A-18

Query Match 3.0%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.11;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 507 CCACCCCTCTTCCAGGGACACCAAGTGGTGTCCAGAGTACATCGAGACGCCGTGCT 566
DB 720 CCACCCCATACCCCTGAACAGAGAGGAGGTTGGACAACTGCTGGAACCAAGCCCATGAT 779
QY 567 CATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGGAAGTGAACCCCT 626
DB 780 GCTGGAGACCAACGACCGCGGCTCTACAAGATCAAGGACACCCACGGCACATCT 839
QY 627 GACCATCTGGTCTACAAGAGAGTTACTTGGGGTCTCAACTCAGCGCTTCTCCCTGGA 686
|||||

DB 840 GACCGCGCGGAGTGGAAACGGCGTGATCCAGCAGATCAAGGCCAACAGCCGCCAGCATCAT 899
QY 687 CAAGCTGGACAGCGCCATCCACTGTGCAACAAAGCGCGTCCAGAAGTACTCTGAAGAATGA 746
DB 900 CGTCGACGACGGCGAGCGCGTGGCCGAGAGAGCGCGTGGCGCCGAGGACTTACGAGAACCC 959
QY 747 TGTGGGCGCAGCCCCCTGCTGCCCGCACACAACATGTGGACCAAGCAGGTTCCAGGA 806
DB 960 CGAGGACAAAGACCCCCAGCCTGACCTGAAGGACGCCCTGAAGCTGAGCTACCCCGACGA 1019
QY 807 GTACCTGGAGCCAGGGCGCGTGGCGCGCTGTGGGCGACGCTCATCTACCGCTCCATGA 866
DB 1020 GATCAAGGAGATCGAGGGCGCTGCTGTACTACAAGAACAAAGCCCATCTACGAGAGCAG 1079
QY 867 GAAGGCCATCGCCCGCAGCGCATGAAGTGGCCCGCAGGACCAACGCTGGAGCCTCGCAAGAACAG 926
DB 1080 GATGACCTATCTAGACGAGAACACCGCCAAAGAGGTGACCAAGCAGCTGAAACGACACCAC 1139
QY 927 CTTTGAGCTTACGGGGCTGACTTCGTCTCTGGGAGGGACTTCAGGCCCTTGGCTGATCGA 986
DB 1140 CGGCAAGTTCAAGGAGCTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
QY 987 GATCA 991
DB 1200 GACCA 1204
RESULT 15
US-08-471-046A-18
Sequence 18, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5866326artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:

```
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CCI695/CIP3/DIV8 - SOLv4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPA(a) 80 kd protein from AB78"
US-08-471-046A-18

Query Match      3.0%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.11;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Qy 507 CCACCTCTTTCCAGGAGCAACAAGTGGGTGCTCCAGAAGTACATCGAGAGCGCGTGCT 566
Db 720 CCACCCCATCACCCTGAACAGAGAGAGTGGACACCTGCTGAACAAAGGCCCATGAT 779
Qy 567 CATCTGTGACACCAAGTTCCAGATCAGACAGTGGTCTCTCGTCACGGACTGGAAACCCCT 626
Db 780 GCTGGAGACCAACAGACCCGCGGCTTCAAGATCAAGGACACCCACCGGCAACATCGT 839
Qy 627 GACCATCTGGTTTACAAAGGAGAGTTACTTGGGTTTCTCAACTCAGCGCTTCTCCCTGGA 686
Db 840 GACCGCGCGAGTGGAAACGCGGTGATCCAGCAGATCAAGGCCAAGACCCGACGATCAT 899
Qy 687 CAAGCTGGAGAGCGCCATCCACCTGTGTCACAAACAGCCGTCAGAAAGTACCTGAAGATGA 746
Db 900 CGTCGACGACGGCGAGCGCGTGGCCGAGAGCGGTGGCCGCCAAGGACTACGAGAACCC 959
Qy 747 TGTGGGCGCGAGCCCTCTGTCGCCGACACACATGTGGACCAACAGGTTCCAGGA 806
Db 960 CGAGGACAAAGCCCCCAGCTGACCTTGAAGGAGCGCCCTGAAGCTGAGCTACCCCGACGA 1019
Qy 807 GTACCTGCAGCGCAGGCGCGTGGCGCGTGTGGGCGAGCGTCTATCTACCCGTCCTATGAA 866
Db 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT 1079
Qy 867 GAAGGCCATCGCCACGCCATGAAGGTGGCCAGGACCAAGTGGAGCCTCGCAAGAACAG 926
Db 1080 GATGACCTATCTAGACGAGAACACCGCAAGGAGGTGACCAAGCAGCTGAACGACACCCAC 1139
Qy 927 CTTTGAGCTCTACGGGCTGACTTCGTCTTGGGAGGGACTTCAGGCCCTGGCTGATCGA 986
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
Qy 987 GATCA 991
Db 1200 GACCA 1204
```

Search completed: September 24, 2005, 11:30:35
Job time : 297.571 secs

This Page Blank (uspio)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1179	72.6	1259	9	AY415398	AY415398 Homo sapi
2	1128.6	69.5	1257	3	AY415399	AY415399 Pan trogl
3	804.4	49.6	2969	3	AK030151	Mus muscu
4	798.2	49.2	2720	3	AK029745	Mus muscu
5	768	47.3	770	7	CR745100	CR745100 CR745100
6	753.8	46.4	1265	9	AY415400	Mus muscu
7	443.6	27.3	775	5	BU937372	AGENCOURT
8	421.4	26.0	4184	3	HSM808322	BU648175 Homo sapi
9	399.6	24.6	777	7	CKS97795	AGENCOURT
10	399	24.6	2520	3	AK080321	Mus muscu
11	378.4	23.3	1023	4	BM080637	AGENCOURT
12	360	22.2	1022	4	BM080516	AGENCOURT
13	321	19.8	693	4	BM717051	UI-B-EJO-
14	316	19.5	861	7	CKS53583	AGENCOURT
15	307.6	19.0	896	6	CD515431	AGENCOURT
16	298.4	18.4	762	6	CD627148	56047432J
17	284.2	17.5	503	7	CKG04006	AGENCOURT
18	258.8	15.9	583	7	CN349423	170004243
19	258.2	15.9	680	6	CB954808	UI-CF-EC1
20	252.8	15.6	514	7	CR746623	CR746623
21	245.4	15.1	320	1	AA144756	mr73d04.1
22	222.4	13.7	429	7	W31827	zb94f04.1
23	219.2	13.5	880	5	BX448465	BX448465
24	213.2	13.1	616	7	CN349424	170006000

```
QY 61 AGCAAGCCAGGACCCAGAGGAGGAGGCGCGGAGCAGCAGCCTGAGCAGCAGGCAAGAT 120
Db 78 AGCAAGCCAGGACCCAGAGGAGGAGGCGCGGAGCAGCAGCCTGAGCAGCAGGCAAGAT 137

QY 121 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 180
Db 138 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 197

QY 181 AAGGTGTCAGAGCCTACCTGGGCGAGCTGAGGAGCAGGACATCGACAGCTCAGCAGAT 240
Db 198 AAGGTGTCAGAGCCTACCTGGGCGAGCTGAGGAGCAGGACATCGACAGCTCAGCAGAT 257

QY 241 GCCGTGAGGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
Db 258 GCCGTGAGGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 317

QY 301 GTTCATGGCGATCTTTCATCTCAATTCAGAAATTAATCTTTTCGAGTGCAGGCTCTG 360
Db 318 GTTCATGGCGATCTTTCATCTCAATTCAGAAATTAATCTTTTCGAGTGCAGGCTCTG 377

QY 361 CTGAATAGAAATCAGCTCTGTGAACCTCTCAGACGACATTCAGCGGCTCCGGAACATCTGG 420
Db 378 CTGAATAGAAATCAGCTCTGTGAACCTCTCAGACGACATTCAGCGGCTCCGGAACATCTGG 437

QY 421 ATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 480
Db 438 ATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 497

QY 481 GAGATCTGAGAGTGGAGCTGAGACACCTCTTTTCAGGGACAAAGTGGGTGGTGC 540
Db 498 GAGATCTGAGAGTGGAGCTGAGACACCTCTTTTCAGGGACAAAGTGGGTGGTGC 557

QY 541 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCGACAGCTG 600
Db 558 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCGACAGCTG 617

QY 601 TTCTCTGTCACGAGCTGGAACCCCTGACCATCTGGTTCACAAAGGAGAGTACTTTGCCG 660
Db 618 TTCTCTGTCACGAGCTGGAACCCCTGACCATCTGGTTCACAAAGGAGAGTACTTTGCCG 677

QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACTGTGCAACAAC 720
Db 678 TTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACTGTGCAACAAC 737

QY 721 GCCGTCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCGCACACAAC 780
Db 738 GCCGTCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCGCACACAAC 797

QY 781 ATGTGACCAAGCACCAGGTTTCCAGGAGTACCTGACGCGCCAGGCGCGTGGCGCGTGTGG 840
Db 798 ATGTGACCAAGCACCAGGTTTCCAGGAGTACCTGACGCGCCAGGCGCGTGGCGCGTGTGG 857

QY 841 GGCAGGCTCATCTACCGGTCCATGAAGAGGCCATCGCCGACGCCATGAAGTGGGCCAG 900
Db 858 GGCAGGCTCATCTACCGGTCCATGAAGAGGCCATCGCCGACGCCATGAAGTGGGCCAG 917

QY 901 GACCACTGAGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTCAGTTCGTTCCTGGG 960
Db 918 GACCACTGAGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTCAGTTCGTTCCTGGG 977

QY 961 AGGGAATTCAGGCGCTTGGTGTGATCGAGATCAATTCAGGCGCCACCATGACCCCGTCCAG 1020
Db 978 AGGGAATTCAGGCGCTTGGTGTGATCGAGATCAATTCAGGCGCCACCATGACCCCGTCCAG 1037

QY 1021 CCGGTCAAGGCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1080
Db 1038 CCGGTCAAGGCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1097

QY 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1140
Db 1098 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1157
```

```
QY 1141 CCATTGAGCGGTCCGACCTCTGCGTGGCGGCGTCAGT 1179
Db 1158 CCATTGAGCGGTCCGACCTCTGCGTGGCGGCGTCAGT 1196

RESULT 2
LOCUS AY415399 1257 bp DNA linear GSS 12-DEC-2003
DEFINITION Pan troglodytes HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY415399
VERSION AY415399.1 GI:39771358
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 1257)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1257)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1257
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1257
/locus_tag="HCM5544"
gene
ORIGIN
Query Match 69.5%; Score 1128.6; DB 9; Length 1257;
Best Local Similarity 96.2%; Pred. No. 1.6e-238;
Matches 1134; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 ATGGCATCCAGCATCTCAAGTGGGTGGTTCAGCCACGAGCTGAGCAGGAGCAGCAG 60
Db 18 ATGGCATCCAGCATCTCAAGTGGGTGGTTCAGCCACGAGCTGAGCAGGAGCAGCAG 77

QY 61 ACAGAGCCAGGACCCAGAGGAGGAGGCGCGGAGGAGCAGCAGCTGAGCAGCAGCAGAT 120
Db 78 AGCAAGCCAGGACCCAGAGGAGGAGGCGCGGAGCAGCAGCTGAGCAGCAGCAGAT 137

QY 121 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 180
Db 138 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 197

QY 181 AAGGTGTCAGGCGCTTACCTGGGCGAGCTGAGGAGCAGGACATCGACAGCTCAGCAGAT 240
Db 198 AAGGTGTCAGGCGCTTACCTGGGCGAGCTGAGGAGCAGGACATCGACAGCTCAGCAGAT 257

QY 241 GCCGTGAGGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
Db 258 GCCGTGAGGACCTCACTGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 317

QY 301 GTTCATGGCGATCTTTCATCTCAATTCAGAAATTAATCTTTTCGAGTGCAGGCTCTG 360
Db 318 GTTCATGGCGATCTTTCATCTCAATTCAGAAATTAATCTTTTCGAGTGCAGGCTCTG 377

QY 361 CTGAATAGAAATCAGCTCTGTGAACCTCTCAGACGACATTCAGCGGCTCCGGAACATCTGG 420
Db 378 CTGAATAGAAATCAGCTCTGTGAACCTCTCAGACGACATTCAGCGGCTCCGGAACATCTGG 437

QY 421 ATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 480
Db 438 ATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 497

QY 481 GAGATCTGAGAGTGGAGCTGAGACACCTCTTTTCAGGGACAAAGTGGGTGGTGC 540
Db 498 GAGATCTGAGAGTGGAGCTGAGACACCTCTTTTCAGGGACAAAGTGGGTGGTGC 557

QY 541 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCGACAGCTG 600
Db 558 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCGACAGCTG 617

QY 601 TTCTCTGTCACGAGCTGGAACCCCTGACCATCTGGTTCACAAAGGAGAGTACTTTGCCG 660
Db 618 TTCTCTGTCACGAGCTGGAACCCCTGACCATCTGGTTCACAAAGGAGAGTACTTTGCCG 677

QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACTGTGCAACAAC 720
Db 678 TTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACTGTGCAACAAC 737

QY 721 GCCGTCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCGCACACAAC 780
Db 738 GCCGTCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCGCACACAAC 797

QY 781 ATGTGACCAAGCACCAGGTTTCCAGGAGTACCTGACGCGCCAGGCGCGTGGCGCGTGTGG 840
Db 798 ATGTGACCAAGCACCAGGTTTCCAGGAGTACCTGACGCGCCAGGCGCGTGGCGCGTGTGG 857

QY 841 GGCAGGCTCATCTACCGGTCCATGAAGAGGCCATCGCCGACGCCATGAAGTGGGCCAG 900
Db 858 GGCAGGCTCATCTACCGGTCCATGAAGAGGCCATCGCCGACGCCATGAAGTGGGCCAG 917

QY 901 GACCACTGAGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTCAGTTCGTTCCTGGG 960
Db 918 GACCACTGAGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTCAGTTCGTTCCTGGG 977

QY 961 AGGGAATTCAGGCGCTTGGTGTGATCGAGATCAATTCAGGCGCCACCATGACCCCGTCCAG 1020
Db 978 AGGGAATTCAGGCGCTTGGTGTGATCGAGATCAATTCAGGCGCCACCATGACCCCGTCCAG 1037

QY 1021 CCGGTCAAGGCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1080
Db 1038 CCGGTCAAGGCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1097

QY 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1140
Db 1098 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1157
```

```
Qy 361 CTGAATAGAATACAGTCTGTGAACCCCTCAGACGGACATTTGACGGCTCCGGAAACATCTGG 420
Db 378 CTGAANNNTTCANNNTNNNNCTNAGACGACATTTGACGGNNCCGGAAACATCTGG 437
Qy 421 ATTATAAGCCCGCGGCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 480
Db 438 ATTATAAGCCCGCGGCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 497
Qy 481 GAGATCTGAGCTGGAGCTGACAGACACCTCTTTCCAGGGACAAAGAGTGGTGGTC 540
Db 498 GAGATCTGAGCTGGAGCTGACAGACACCTCTTTCCAGGGACAAAGAGTGGTGGTC 557
Qy 541 CAGAAGTACATCGAGACGCGCTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
Db 558 CAGAAGTACATCGAGACGCGCTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 617
Qy 601 TTCCTCGTCCAGGACTGGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTCGGG 660
Db 618 TTCCTCGTCCAGGACTGGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTCGGG 677
Qy 661 TTCCTCAACTGAGCTTCTCTCGGACAGCTGACAGCGGCATCTGACCTGTGCAACAAC 720
Db 678 TTCCTCAACAGAGCTTCTCTCGGACAGCTGACAGCGGCATCTGACCTGTGCAACAAC 737
Qy 721 GCCGTCCAGAGTACCTGGAAGATGATGTGGCCGCGACGCCCTGCTGCCGCGACACAAC 780
Db 738 GCCGTCCAGAGTACCTGGAAGATGATGTGGCCGCGACGCCCTGCTGCCGCGACACAAC 797
Qy 781 ATGTGGACACAGCAGGTTCCAGGAGTACTGTGAGCGCCAGGCGCGCTGGCGCTGG 840
Db 798 ATGTGGACACAGCAGGTTCCAGGAGTACTGTGAGCGCCAGGCGCGCTGGCGCTGG 857
Qy 841 GGCAGCTCATCTTACCGTTCATGAAGAGCCATCGCCCAACGCGCATGAAGGTGGCCAG 900
Db 858 GGCNNNGTCATCTACCGCTTCATGAAGAGCCATCGCCCAACGCGCATGAAGGTGGCCAG 917
Qy 901 GACACGTGGAGCTCCAGACAGCTTTGAGTCTACGGGGTGACTCGCTCTGGTGG 960
Db 918 GACACGTGGAGCTCCAGACAGCTTTGAGTCTACGGGGTGACTCGCTCTGGTGG 977
Qy 961 AGGACTTCAGGCGCTGGCTGATCGAGTCAATTCAGAGCCCAACCATGACCCGCTCCAG 1020
Db 978 AGGACTTCAGGCGCTGGCTGATCGAGTCAATTCAGAGCCCAACCATGACCCGCTCCAG 1037
Qy 1021 CCGGTACGCGCCAGCTGTGTGCAAGGTGACAGGACACCATCAAGGTGGCGGTGGAC 1080
Db 1038 CCGGTACGCGCCAGCTGTGTGCAAGGTGACAGGACACCATCAAGGTGGCGGTGGAC 1097
Qy 1081 CGCAGCTGTGACATCGGCACTTCGAGTCTGTGGAGGACGCGGTGGTGGAGCCGCC 1140
Db 1098 CGCAGCTGTGACATCGGCACTTCGAGTCTGTGGAGGACGCGGTGGTGGAGCCGCC 1157
Qy 1141 CCATTACGCGGTCGACCTCTGGTGGCGGCGTCAGT 1179
Db 1158 CCATTACGCGGTCGACCTCTGGTGGCGGCGTCAGT 1196

RESULT 3
AK030151
LOCUS AK030151 2969 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4933401B17 product:hypothetical Tubulin-tyrosine
ligase containing protein, full insert sequence.
ACCESSION AK030151
VERSION AK030151.1 GI:26326134
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
```

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

TITLE

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

20499374
PUBMED
11042159

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

20530913
MEDLINE
11076861

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL

Nature 409, 685-690 (2001)

REFERENCE

The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL

Nature 420, 563-573 (2002)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N.,
Okazaki, Y., Saito, K., Saito, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/

FEATURES

Location/Qualifiers
1..2969
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4933401B17"
/db_xref="taxon:10090"
/clone="4933401B17"
/sex="male"

/tissue type="testis"
/clone lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"
490_-2835
/note="unnamed protein product; hypothetical
tubulin-tyrosine ligase containing protein
(InterPro|IPR004344, evidence: InterPro)
putative"
/codon start=1
/protein_id="BAC26811.1"
/db_xref="GI:26326135"
/translation="MENGGRKKLUSSTLSDDGHHKKNLQGIQDILSSSPKLDRYKIA
RQTEKAIKERKIFSIYGHYPVIRLRRKGGWKEKFFPKALQNLGSEDKSATKE
NOIALERFDDIHVMSRLVKNIEPYLLMTIKEDVDVYHSLTCDQMLNHGKTASFTT
KIGLCINMRSLPWWQANPNTFPFCKGLCTSEKOEFLDDPRFTVAASILKVVVLHQ
NYCSKVKGSKKEEAKNPSPKDPENPDLPKLSGOVVDTCVKCAOYLGLHE
DIDVSASFEALSEEWNDLTQQYILLVHGNASITDSKSYFACQALLSKISVNPQT
EIDGIRNIWIKPAASGRDVCMDRVENILSVLAADSQTTKDNKWVQKIETPML
IYDKFDIRQWFLVTDWNPITWFKESYLRFSTQRFSLDKLSAHLNCSNIOQRLLK
NDRKSPFLPCHNMWTSFQBYLQKRGRTWGSIIYPSMKRAVTNMRVAQDHVEA
RKNSELYGADFILGRDFKFWLIEINSSPTMHPSPTVTAQLCAQOVEDTIKVVDRLK
DRNCDIGNFELLWRQPAVELPPNGSDLCVEGIVSVKAKKQMPPIASVGLSESLDAP
PKVRSARALMETVIRPRTVROQDWKEBAKVLSTTWSMPVMDAEVGRGAKPIYAEV
NDIQVDNKGSHGSIYTRVQSSKVPGLTSAQHPALFAQTMKPTQMTSSPPPTASGNH
RDSPPFCPIVFEELWLHPNSQRRPSSCILQSRQAQWIRGIP"

ORIGIN

Query Match 49.6%; Score 804.4; DB 3; Length 2969;
Best Local Similarity 74.8%; Pred. No. 7.5e-167;
Matches 1057; Conservative 0; Mismatches 336; Indels 21; Gaps 3;

QY 2 TGGCATCCAGCATCCTCAAGTGGGTGGTCAGCCACCAGAGCTGCGAGGAGCAGCAGAA 61
DB 1106 TGGCTGCCAGCATCCTCAAGTGGGTAGTCTCTTCAACAGAAATTAC-----TGCAGCAAAAG 1159

QY 62 GCAAGCCAGGACACGAGGAGAGGAGCCGGGAGCAGCGACTGAGCAGGAGCAAGATG 121
DB 1160 TCAAGGCGAAGATAAGAGGAGGAGGCCAAGAACAGTGCACCCAGCCCCCAAGAAAGATC 1219

QY 122 CTCGAAATCTGAGGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGGGACATCGCGTGCA 181
DB 1220 CTGAAATCTGACCTCAAGCTACCAAGCTCTCGGGACAGGTTGTAGACACAGCTTGCA 1279

QY 182 AGGTGTGCCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACAGTCAGCAGATG 241
DB 1280 AGGTGTGCCAGGCTTACCTGGGACAGCTAGAGCATGAAGACATAGAGTATCGGAGGCCA 1339

QY 242 CCGTGGAGGACCTACTAGGCGCAGTGGGAGGACCTGACCGAGCATGACTACTCCTCG 301
DB 1340 GCACCGAGGCTTGTCTGAGGAGGAATGGAAATGACCTGCACAGCAGTACTATCTACTGG 1399

QY 302 TTCATGGCGATGCTTTTCATCTCAAAATTCAGAAATTTACTTTTCGAGTGCAGGCTCTGC 361
DB 1400 TTCATGGCAATGCTTTCATCCTGATTCGAAAGTTACTTTTGGCAATGCCAGCTCTGC 1459

QY 362 TGAATAGAATCAAGTCTGTGAACCTCCAGACGACATTCACGGGCTCCGGAACATCTGGA 421
DB 1460 TGAGTAAGATCAGTCTGTGAACCTCCAGACAGATCGATGGGATCCGGAACATCTGGA 1519

QY 422 TTATAAGCCCGGGCAGTCCCGGGGCGGAGACATAGTGTGCATGAGCCGCTGTGGAGG 481
DB 1520 TCATAAAGCCTGAGCAGCAAGTCCCGGGTGCAGATATTTGTGTATGGACCGCTGTGAGA 1579

QY 482 AGATCTGGAGCTGGCAGCTGCAGACACCTCTTTCCAGGGACAACAAGTGGTGGTCC 541
DB 1580 ACATCTGTAGTCTGTGTGGCGGAGACAGCAGACCAAGAGCAACAATGGTGGTCTC 1639

QY 542 AGAAGTACATCGAGACCGCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGT 601
DB 1640 AGAAGTACATCGAGACCCCAATGCTCATCTACGACACCAAGTTTTCGACATCAGACAGTGT 1699

QY 602 TCCTCGTACGAGTACGAAACCCCTGACCATCTGGTTCTTACAGGAGAGTACTTCGCGT 661

DB 1700 TCTCGTCAAGACTGGAATCCCTAAACCATCTGTTTCTACAAAGAAAGACTACTCTCGCT 1759
QY 662 TCTCAACTCAGCGCTTCTCCCTGGCAAGACTGGAAGCGCCATCCACCTGTGTGCAACAAAG 721
DB 1760 TCTCCACACAGCGCTTCTCCCTGGCAAGACTGGAAGCGCCATCCACCTGTGTAACT 1819
QY 722 CCGTCCAGAAAGTACTTGAAGAATGATGTGGGCGGAGCCCCCTGTGTGCCCGCACAAACA 781
DB 1820 CCATCCAGAGGCGCTCAAAAATGATAAAGAGCGAGTCCGCTGTCTTGTCTATAACA 1879
QY 782 TGTGACACGACACCGAGTTCCAGGAGTACTCTCAGCGCCAGCGGCGGTGGCGCGTGGG 841
DB 1880 TGTGACACGAGTCTGTCTTCCAGGAGTACTCTCAGGAGGCGGCGGAGAGGAACGTGG 1939
QY 842 GCAGCGTCACTTACCGCTCCATGAAGAAGCCATCGCCACGCCATGAAGGTGGGCCAGG 901
DB 1940 GTAGCATCATCTACCGCTCTATGAAAAGAGCTGTACCATGCCATGAGGTAGCCAGG 1999
QY 902 ACCAGTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTCTTGGGA 961
DB 2000 ACCAGTGAAGCCCGTAAGAACAGCTTTCGAGCTCTACGAGCGGACTTCATCTCTGGGC 2059
QY 962 GGGACTTCAGGCGCTGGCTGATCGAGATCAATCCAGCGCCACCATGCAACCGTCCACGC 1021
DB 2060 GAGACTTCAAGCCATGGCTTATCGAGATCACTCCAGCGCCCATGCAACCCCTCCACTC 2119
QY 1022 CGGTACCGGCCAGCTGTGTGCACAGGTGCAGAGGACACCATCAAGGTGGCG----- 1075
DB 2120 CGGTACAGCTCAGCTCTGTGCCAGGTGCAGAGGACACCATCAAGGTGGTGGATC 2179
QY 1076 -----TGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGAGCGCGTGG 1129
DB 2180 GCAAACTGGACCGGAACTGTGACATCGGCAACTTTTGTGAGCTTCTGTGGCGGCGCTGCCG 2239
QY 1130 TTGAGCGCGCCCATTCAGCGGCTCGACCTCTGCGTGGCGGCGCTGAGTGTGAGAGAG 1189
DB 2240 TGGAGTGCACCGTTCACCGGCTGTGACCTCTGTGTGGAAGGTATCAGCGTGAAGAAAG 2299
QY 1190 CCAGGAGGAGGAGTGTGCGCGCTCTGCAACCTCAAGCGCTCGGCTCTGTTGGAGCGCG 1249
DB 2300 CCAAGAGCAATGCGGCCATTTGCGAGGTGCGGCTCTCAGAACTACTTTAGATGC-- 2357
QY 1250 AGCCGCTGAAGGACAGGGGCCCTCGGCCATGCCAGACCTGCCAGGAGACCCCATCAC 1309
DB 2358 -TCCACCCAAAGTGGCGAGTGCCTGGGCGCTGATGAAACAGTGTATCAGCGCTCCAGAA 2416
QY 1310 CAGCTCTCCAGCGGAGCTTGGGAGCTGAAGAGAGAGAGGGGCTCCCTGGCGCTTGTCTGG 1369
DB 2417 CGACCGTGGCGAGGACTGGAACCGGAGAGCAAGGTACTTTCAACCACTTGTGTCCA 2476
QY 1370 CACCTTTAAGGGGGCGAGCGGAGAGCGGTGGAGC 1403
DB 2477 TGCCAGTAATGATGCAGAGGTGAGAGGTGAGG 2510

RESULT 4

AK029745
LOCUS 2720 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:490524K07 product:hypothetical Tubulin-tyrosine ligase containing protein, full insert sequence.
ACCESSION AK029745
VERSION AK029745.1 GI:26325681
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
MEDLINE 99279253

QY 722 CCGTCCAGAGTACCTGAAGAATGATGTGGGCGCAGCCCCCTGTGTCGCCGCACACA 781
Db 1823 CCATCCAGAGCGCCTCAAAATGATAAAGAGCGAGTCGCTGCTACCTGTGCTATAACA 1882
QY 782 TGTGGACCAAGCAGGTTCCAGAGTACTGTCAGCGCCAGGCGCTGGCGCGTGTGG 841
Db 1883 TGTGGACCAAGCAGTCTGTTTCCAGAGTACTGTCAGAGAGGCGCGAGGAGAACGTGG 1942
QY 842 GCAGCGTCTATCTACCGTCCATGAAGAGCCATCGCCACGCGCATGAAGGTGGCCAGG 901
Db 1943 GTAGCATCATCTACCGTCTATGAAGAAGCTGTACCAATGCCATGAGGTAGCCAGG 2002
QY 902 ACCAGTGGAGCCTCCAAAGAACAGCTTTAGCTCTACGGGGTGTGCTTCCTTTGGA 961
Db 2003 ACCAGTGAAGCCGTAAGAACAGCTTCAGCTCTACGAGCGGAGCTTCATCTGGGGC 2062
QY 962 GGGACTTCAGGCGCTCGGTGATCGAGATCAATTCAGCGCCACCATGCAACCGTCCAGC 1021
Db 2063 GAGACTTCAAGCCATGGCTTATCGAGATCAACTCCAGCGCCACCATGCAACCGTCCACTC 2122
QY 1022 CGGTACGGGCCAGCTGTGTGACAGGTGCGAGGAGACACCATCAAGGTGGCG----- 1075
Db 2123 CGGTACAGCTAGCTCTGTGCCAGGTGCGAGGAGACACCATCAAGGTGGTGGATC 2182
QY 1076 -----TGGACCGCAGCTGTGACATCGGAACATTCGAGCTCTGTGGAGCGACCGGTGG 1129
Db 2183 GCAAACTGGACCGAAACTGTGACATCGGAACATTTGAGCTCTGTGGCGGAGCCTGCCG 2242
QY 1130 TTGAGCGGCGCCCATTTACGGGGTCCGACTCTGCTGGCGGGCGTCACTGTGAGAGAG 1189
Db 2243 TGGAGTGGCCACCGTTCAACGGGTCTGACCTCTGTGTGGAAGGTATTCAGCGTGAAGAA 2302
QY 1190 CCAGGAGGAGGTGCTGCCGTCTGCAACTCAAGCGCTCGGCTCGCTGTTGGAGCGCG 1249
Db 2303 CCAGAGCAATATGCCGCCCATTTGCCAGGTGCGGCTCTCAGATCATTAGATGC-- 2360
QY 1250 AGCGCTGAAGGCA CGGGGCCCTCGGCCATGTCAGACCCCTGCCAGGAGACCCCATCAC 1309
Db 2361 -TCCACCCAAAGTGGGAGTGGCCGGCCCTGATGAAACAGTGATCAGCGCTCCAAAGAA 2419
QY 1310 CAGCTCTCAGCGGAGCTTGGACTGAAGAGAGAGAGGGGCTCCCGCTGGCTTGGTGG 1369
Db 2420 CGACCGTGGCGGAGGACTGGAACCGGGAAGAACAAAGGGCTTCGCTGTGATCTCOAGC 2479
QY 1370 CACCCCTTAAGGGGGGCGAGAGCGGTGGAGCGGTGGAGCCGACAGCCACCCGACCAAAGCTG 1429
Db 2480 GCGTGCCTAGTTGTGTGTGGCTTCGCGCGCGCCCTGGCGCTGCAAGCTCTCCCT 2539
QY 1430 CTGGGAAGGTGGAGCTCCCGGCTGCCCCCTGTGTCG 1464
Db 2540 GCGCCTTGATGCGCGAGGGCGCCTCTGTGTGCTGC 2574

RESULT 5
CR745100 770 bp mRNA linear EST 30-AUG-2004
LOCUS CR745100 Soares testis NHT Homo sapiens cDNA clone IMAGp97102270 ;
DEFINITION IMAGE:1645239 5', mRNA sequence.
ACCESSION CR745100
VERSION CR745100.1 GI:51667573
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES 1 (bases 1 to 770)
JOURNAL Ebert,L., Heil,O., Hennig,S., Korn,B., Neubert,P., Partsch,E.,
Peters,M., Radelof,U. and Schneider,D.
COMMENT I.M.A.G.E. cDNA Clone Collection
Unpublished (2004)
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany

Email: www.rzpd.de
RZPD; IMAGp97102270.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTACACAGAAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..770
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp97102270 ; IMAGE:1645239"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaudo."

ORIGIN

Query Match 47.3%; Score 768; DB 7; Length 770;
Best Local Similarity 99.7%; Pred. No. 6.9e-159;
Matches 768; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 251 ACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCG 310
Db 1 ACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCG 60
QY 311 ATGCTTTTCATCTCCAATTCAGAAATTAATTTTCGACGTGCGAGGCTCTGCTGAATAGAA 370
Db 61 ATGCTTTTCATCTCCAATTCAGAAATTAATTTTCGACGTGCGAGGCTCTGCTGAATAGAA 120
QY 371 TCACGTCTGTGAACCCCTCAGACGGACATTCAGCGGCTCGGGAACATCTGGATTATAAAGC 430
Db 121 TCACGTCTGTGAACCCCTCAGACGGACATTCAGCGGCTCGGGAACATCTGGATTATAAAGC 180
QY 431 CCGCGGCCAAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCGTGTGGAGGAGATCCTGG 490
Db 181 CCGCGGCCAAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCGTGTGGAGGAGATCCTGG 240
QY 491 AGCTGGCAGCTGCAGACCAACCCCTTTTCAGGGAACAAGTGGGTGTCAGAGTACA 550
Db 241 AGCTGGCAGCTGCAGACCAACCCCTTTTCAGGGAACAAGTGGGTGTCAGAGTACA 300
QY 551 TCGAGACGCGCGTGTCTATCTGTGACACCAAGTTGACATCAGACAGTGGTTCTCTCGTCA 610
Db 301 TCGAGACGCGCGTGTCTATCTGTGACACCAAGTTGACATCAGACAGTGGTTCTCTCGTCA 360
QY 611 CCGACTGTGAACCCCTCGTACCATCTGGTCTTACAAGGAGAGTTACTTTCGGTTCTCAACTC 670
Db 361 CCGACTGTGAACCCCTCGTACCATCTGGTCTTACAAGGAGAGTTACTTTCGGTTCTCAACTC 420
QY 671 AGCGCTTCTCTCGTGAACAAGCTGGAGCGCCATTCACCTGTGCAACAACCGCGTCCAGA 730
Db 421 AGCGCTTCTCTCGTGAACAAGCTGGAGCGCCATTCACCTGTGCAACAACCGCGTCCAGA 480
QY 731 AGTACTGAAGATGATGTGGCGCGCCCGCTCTGCCCGCACACACATGTGGACCA 790
Db 481 AGTACTGAAGATGATGTGGCGCGCCCGCTCTGCCCGCACACACATGTGGACCA 540

```
QY 791 GCACCGGTTCCAGAGTACTCTGACGCGCCAGGCGCGTGGCGCCGCTGTGGGCGACGCTCA 850
Db 541 GCACCGGTTCCAGAGTACTCTGACGCGCCAGGCGCGTGGCGCCGCTGTGGGCGACGCTCA 600
QY 851 TCTACCGGTTCCATGAAGAGCCCATCGCCACGCGCATGAAGTGGCCGCCAGGACCAAGCTGG 910
Db 601 TCTACCGGTTCCATGAAGAGCCCATCGCCACGCGCATGAAGTGGCCGCCAGGACCAAGCTGG 660
QY 911 AGCCTCGCAAGAACAGCTTTGAGCTCTACCGGGGCTGACTTGGTCTTTGGGAGGAGCTTCA 970
Db 661 AGCCTCGCAAGAACAGCTTTGAGCTCTACCGGGGCTGACTTGGTCTTTGGGAGGAGCTTCA 720
QY 971 GGCCTCGGCTGATCGAGATCAATTCAGCCCCCAGCCCATGACCCCTCCACG 1020
Db 721 NGCCCTGGCTGATCGAGATCAATTCAGCCCCCAGCCCATGACCCCTCCACG 770

RESULT 6
LOCUS AY415400 1265 bp DNA linear GSS 12-DEC-2003
DEFINITION Mus musculus HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY415400
VERSION AY415400.1 GI:39771359
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1265)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferred nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
JOURNAL PUBMED 14671302
REFERENCE 2 (bases 1 to 1265)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source location/Qualifiers
1..1265
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1265
/locus_tag="HCM5544"

gene
ORIGIN
Query Match 46.4%; Score 753.8; DB 9; Length 1265;
Best Local Similarity 78.6%; Pred. No. 1e-155;
Matches 934; Conservative 0; Mismatches 237; Indels 18; Gaps 2;

QY 2 TGGCATCCAGCATCCTCAAGTGGGTGTGACGCCACCGAGCTCGAGCAGGACGACGAGAA 61
Db 19 TGGCTGCCAGCATCCTCAAGTGGGTGTGACGCCACCGAGCTCGAGCAGGACGACGAGAA 72
QY 62 GCAAGCCAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121
Db 73 TCAAGGCGCAAGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 132
QY 122 CTGAAATGCTGAGGCAAGCTCAGGCGCTCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 181
Db 133 CTGAAATGCTGAGGCAAGCTCAGGCGCTCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 192
```

```
QY 182 AGTGTGCCAGGCGCTACTCTGGGGCAGCTGGAGGAGCATGAGGAGCATCGACACAGTCTCAGAGATG 241
Db 193 AGTGTGCCAGGCGCTACTCTGGGGCAGCTGGAGGAGCATGAGGAGCATCGACACAGTCTCAGAGATG 252
QY 242 CCGTGGAGGACCTCAGTGGAGGCGGAGTGGGAGGAGCCTGACCCAGCAGTACTACTCCCTCG 301
Db 253 GCACCGAGGCGCTGTCTGAGGAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 312
QY 302 TTCAATGAGATGAGTCTTTCATCTCCAAATTCAGAAATTAATTTTCAGAGTCCGAGGCTCTGCG 361
Db 313 TTCAATGAGATGAGTCTTTCATCTCCAAATTCAGAAATTAATTTTCAGAGTCCGAGGCTCTGCG 372
QY 362 TGAATGAGATGAGTCTTTCATCTCCAAATTCAGAAATTAATTTTCAGAGTCCGAGGCTCTGCG 421
Db 373 TGAATGAGATGAGTCTTTCATCTCCAAATTCAGAAATTAATTTTCAGAGTCCGAGGCTCTGCG 432
QY 422 TTATAAAGCCCGCGGCGCAAGTCCCGGGCGGAGACATAGTGTGATGATGATGATGATGATGATGAT 481
Db 433 TCATTAAGCCCTGAGCCGAGTCCCGGGTGGAGTATTTGTGTATGATGATGATGATGATGATGATGAT 492
QY 482 AGATCTCTGGAGTGGCAGCTGACAGCACCCTCTTTCCAGGGACCAACAAGTGGGTGGTTC 541
Db 493 ACATCTCTGAGTCTGGTGGCGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 552
QY 542 AGAAGTACATCGAGAGCGCGCTGCTCATCTGTGACACCAAGTTTCGACATCGACAGTGGT 601
Db 553 AGAAGTACATCGAGAGCGCGCTGCTCATCTGTGACACCAAGTTTCGACATCGACAGTGGT 612
QY 602 TCCTCGTCAAGGAGTGGAGCCCTGACCATCTGTTTACAGAGGAGTACTTTCGGT 661
Db 613 TCCTCGTCAAGGAGTGGAGCCCTGACCATCTGTTTACAGAGGAGTACTTTCGGT 672
QY 662 TCCTCAACTCAGCGCTTCTCCCTGGAGCAAGCTGGAGCGGCATCCACTGTGCAACAACG 721
Db 673 TCCTCCACACAGCGCTTCTCCCTGGAGCAAGCTGGAGCGGCATCCACTGTGCAACAACG 732
QY 722 CGCTCCAGAGTACTTGAAGATGATGTGGCGCGCAGCGCCCTGTCTGCCCGCACACAACA 781
Db 733 CCATCCAGAGCGCTTCAAAAATGATAAAGAGCGCAGTCCGCTGCTACCTTGTGTAACA 792
QY 782 TGTGGAGCAGCAGCAGGTTCCAGAGTACCTGAGCGCGCAGCGCGCGCGCGCGCGCGCGCGCG 841
Db 793 TGTGGAGCAGCAGCAGTCTGTTCCAGAGTACCTGAGAGAGGCGCGCGCGCGCGCGCGCGCGCG 852
QY 842 GCAGCGCTACTTACCGCTCCATGAAGAGCCCATCGCCACGCGCATGAAGTGGCGCCAGG 901
Db 853 GTAGCATCATCTACCGCTCTATGAAGAGCTGTCAACATGCGATGAGGAGTGGAGTGGAGTGGAG 912
QY 902 ACCAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGTGTGACTTCCTCTCTGTTGGGA 961
Db 913 ACCAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGTGTGACTTCCTCTCTGTTGGGA 972
QY 962 GGGACTTTCAGGCGCTGCTGATCGAGTCAATTTCCAGCGCCCGCAGCGCGCGCGCGCGCGCGCG 1021
Db 973 GAGACTTTCAGCGCTTATCGAGTCAATTTCCAGCGCCCGCAGCGCGCGCGCGCGCGCGCGCG 1032
QY 1022 CGGTACCGCGCCAGCTGTGTGCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1075
Db 1033 CGGTACAGCTCAGCTCTGTGCCAGGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1092
QY 1076 -----TGAACCGCAGCTGTGACATCGGCAACTTTCGAGTCTCTGTGAGGAGGAGGAGGAGGAG 1129
Db 1093 GCAAACTGGACCGGAACTGTGACATCGGCAACTTTCGAGTCTCTGTGAGGAGGAGGAGGAGGAG 1152
QY 1130 TTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1178
Db 1153 TGGAGCTGCCACCGTTCAACGGGCTCTGACCTCTGTGTGGAAGGATATCAG 1201

RESULT 7
LOCUS BU937372 775 bp mRNA linear EST 18-OCT-2002
```

DEFINITION AGENCOURT_10519180 NIH_MGC_169 Mus musculus cDNA clone
IMAGE:6705256 5', mRNA sequence.
ACCESSION BU937372
VERSION BU937372.1 GI:24126191
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1. (bases 1 to 775)
NIH-MGC <http://mgc.nci.nih.gov/>
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabp-r@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM3002 row: g column: 16
High quality sequence stop: 569.
Location/Qualifiers
1. .775
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6705256"
/lab_host="DH10B (Ti-phrase-resistant)"
/clone_lib="NIH MGC 169"
/note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI
(ggccatcgcc); Site 2: SfiI (ggccctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGAGTGGCCATTCAGCGCGG-3' and
5'-ATTCTAGAGCGGAGCGCGGCACATG-dt(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 27.3%; Score 443.6; DB 5; Length 775;
Best Local Similarity 74.5%; Pred. No. 3.1e-87;
Matches 576; Conservative 0; Mismatches 185; Indels 12; Gaps 1;
QY 586 GACATCAGACAGTGGTTCTCGTCACGAGCTGGAACCCCTGACCATCTGGTTCTACAAG 645
Db 1 GGCATCAGACAGTGGTTCTCGTCACAGACTGGAATCCCTTAACCATCTGGTTCTACAAA 60
QY 646 GAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCGACGAAATGATGTGGCGCGCAGCGCCCTG 705
Db 61 GAAAGCTACTCTCGCTTCTCCACACAGCGCTTCTCCCTGGACAAAGTGGACAGCGGCATC 120
QY 706 CACCTGTGCAACAAACCCCTGCAGAAAGTACCTGGAAGATGATGTGGCGCGCAGCGCCCTG 765
Db 121 CACCTGTGTAACTCCATCCATCAGAGCGCGCTCAAAATGATAAGAGCGCATCGCTG 180
QY 766 CTGCCCGCACACAACATGTGGACACAGACACAGGTTCCAGAGTACTGTGAGCGCCAGGCG 825
Db 181 CTACCTTGTATACATATGTGGACACAGCACTCGTTTTCCAGAGTACCTTCGAGAAGAGGGGC 240
QY 826 CGTGGCGCGTGTGGGGAGCGTCATCTACCGTCCATGAAGAGCCCATCGCCACGCC 885
Db 241 CGAGGAGGAACGTGGGGTAGCATCATCTACCGCTTATGAAAGAGCTGTGACCAATGCC 300
QY 886 ATGAAGTGGCCCGAGGACCAACGTGGAGCGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCT 945
Db 301 ATGAGGGTAGCCAGGACCAACGTAGAGCCCGTAAGAACAGCTTTCAGCTCTACGAGGCC 360
QY 946 GACTTCGTCTTGGGAGGGAGCTTTCAGGCGCTGCTGATCGAGATCAATTCAGGCCCCACC 1005

Db 361 GACTTCATCTCTGNGCGAGACTTCAAGCCATGGCTTATCGAGATCAACTCCAGCCCCACC 420
QY 1006 ATGCACCCGCTCAGCGCGGTTCAGGGCCCGAGCTGTGTGCACAGGTGCAGGAGGACACCATC 1065
Db 421 ATGCACCCCTCCACTCCCGTCAAGCTCAGCTGTGTGCCAGGTGCAGGAGGACACCATC 480
QY 1066 AAGGTGGCG------TGACCCAGCTGTGACATCGGCAACTTCGAGTCTCTG 1113
Db 481 AAGGTGGTGGTGGATCGCAAACTGGACCGAAACTGTGACATCGGCAACTTTGAGCTTCTG 540
QY 1114 TGGAGCAGCCGTTGAGTGGTTCAGCGCGCCCATTCAGCGGGTCCGACCTCTCGTGGCGGGC 1173
Db 541 TGGCGGACGCTCGCGTGGAGTGGCCACCGTTCAACGGGTCTGACCTCTGTGTGGAAGGT 600
QY 1174 GTCAGTGTGAGGAGCGCAGGAGGAGGTGCTGCCCGTCTGCAACCTCAAGGCTCTCGGCC 1233
Db 601 ATCAGCGTGAAGAAGCCCAAGAGCAATATGCCGCCATTTGCCAGCGTTCGGCTCTCAGAA 660
QY 1234 TCGCTGTTCGAGCGCGCAGCGCTGAGGACACGCGGGCCCTCGGCCATGCCAGACCTGCC 1293
Db 661 TCACTCTTAGATGCTCCACCCAAAGTGGCGAGTGGCCCGGCCCTGTGATGAAACAGTGA 720
QY 1294 CAGGAGACCCCATCACCAGCTCTCCAGCGGAGCTTGGGACTGAGGAAGAGAA 1346
Db 721 TCAGGCTTCAAGAACGACACCGCTGCGGCGAGACTGGGAAACGCGGGAAGAAGCA 773
RESULT 8
HSM808322 Homo sapiens mRNA; cDNA DKFZp686D076 (from clone DKFZp686D076).
LOCUS BX648175
ACCESSION BX648175.1 GI:34367334
VERSION
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 4184)
AUTHORS Ansonge, W., Krieger, S., Regiert, T., Rittmueller, C., Schwager, B.,
Mewes, H.W., Weill, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
CONSRMT The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686D076) is available at the RZPD Deutsches
Resourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
<http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686D076>
Further information about the clone and the sequencing project is
available at <http://mips.gsf.de/projects/cdna/>.
FEATURES
Location/Qualifiers
1. .4184
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp686D076Q"
/db_xref="taxon:9606"
/clone="DKFZp686D076"
/tissue_type="cervix"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
/note="hypothetical protein"
1. .4184
/gene="DKFZp686D076"
2529. .3587
gene
CDS

Db	3411	TGTGACACAGGACCTTTGAGCTCATCTATAAGACCCCGT	3451
RESULT 9			
CKS97795			
LOCUS	CKS97795	777 bp	mRNA
DEFINITION	AGENCOCURT 17837841 NIH MGC 238 Rattus norvegicus cDNA clone		linear
	IMAGE:7135519 5', mRNA sequence.		
ACCESSION	CKS97795		
VERSION	CKS97795.1	GI:41110910	
KEYWORDS	EST.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 777)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgabs-r@mail.nih.gov Tissue Procurement: Howard Jacobs cDNA Library Preparation: Express Genomics cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM15042 row: 0 column: 05 High quality sequence stop: 693. Location/Qualifiers		
FEATURES	1..777		
source	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10116"		
	/clone="IMAGE:7135519"		
	/tissue_type="testis, pooled"		
	/lab_host="DH10B Tona"		
	/clone_lib="NIH MGC 238"		
	/note="Organ: testis; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTAGTCTAGATCGGAGCGCGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >1.4kb resulted in an average insert size of 1.9 kb. This primary library is normalized (non-normalized primary library is NIH MGC 237) and was constructed by Express Genomics (Frederick, MD)"		
ORIGIN			
Query Match	24.6%	Score 399.6;	DB 7; Length 777;
Best Local Similarity	77.3%	Pred. No. 1.6e-77;	

Matches	513;	Conservative	0;	Mismatches	144;	Indels	7;	Gaps	2
---------	------	--------------	----	------------	------	--------	----	------	---

Qy	2	TGGCATCCAGCATCCTCAAGTGGGTGGTGTCACGCCACGAGCTGCAGCAGGACGACGACAA	61
Dd	43	TGGCTGCCAGCATCCTCAAGTGGGTAGTCCATCACCGAATTAC-----TGCAACAGAG	96
Qy	62	GCAAGCCAGGGAACAGAGGAGGAGCGCCGGGAGCAGCGACTGAGCAGGACGGAAGAATG	121
Dd	97	CCAAGGGCAAGAGTAAGAAGAGGAGGCGCAAGAAATGGTGACCCACGCCCCAAGAAGATC	156
Qy	122	CTGAAATGCTGAGGCAAGCTCAGGGGCTCCCGGGCAGCTTTGGACATCGCGTGCA	181
Dd	157	CTGATAATCTGACCTCAAGTTACCAAGCCCTCTCAGGACAGGTTGTAGACACAGCATGCA	216
Qy	182	AGGTGTGCAGGCCCTACTCTGGGGCAGCTGGAGCATGAGGACATCGACAACGTCACAGATG	241
Dd	217	AGGTGTGCAGGCCCTACTCTGGGACAGCTGGAGCATGAGGACATCGACTGTACACGCGA	276
Qy	242	CCGTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACCAGCAGTACTTCTCCCTCG	301
Dd	277	CGCCGAGGGCCTTGTCGAGGAGGAATGGAACGACCTGACACAAATACTTACTTGTCTGG	336
Qy	302	TTCAATGCGGATGCTTTCATCTCCAAATTCAGAAAATTAATTTTCGAGTGCACAGCTCTGC	361
Dd	337	TTCAATGCGGATGCTTTCATCACGGAATTCGAGAAGTTACTTTTGGCGAGTGCACAGCTCTGC	396
Qy	362	TGAATAGAATCACGCTCTGTGAACCTCAGACGCGACATTTGACGGGCTCCGGAACATCTGGA	421
Dd	397	TGAGTAGATCAGCTCTGTGNACCCCAGACCGAGATTTGACGGGATAGGAGACATCTGGA	456
Qy	422	TTATPAAAGCCCGGCCAACGATCCCGGGCCGAGACATAGTGTGATGACCCGCTGGAGG	481
Dd	457	TCATTAAGCCTGCAGCCAAAGTCCCGGGCCGAGATATTGTGTGATGACCCGCTGGAGA	516
Qy	482	AGATCTTGGAGCTGGAGCTGCAGACCACTCTTTTCAGGGACAAACAGTGGGTGTGTC	541
Dd	517	ACATCTTGGATCTGGTGCCACACAGACGCGACACCAAGGAGCAACAAATTTGGTGTGTC	576
Qy	542	AGAAGTACATCGACAGCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGT	601
Dd	577	AGAAGTACATCGACAGCCGATGCTCATCTATGACACCAAGTTTCGACATCAGACAGTGGT	636
Qy	602	TCCTCGTACGGAAGTGGAAACCCCTGACCATCTGGTTCTACAGGAGAGTACTTGGCGGT	661
Dd	637	TCCTTGTACAGACTGGAATCCCT--AACATCTGTTCTACAGAGAGACTACCTCGGTC	695
Qy	662	TCTC 665	
Dd	696	TCAC 699	

RESULT	10
AK080321	
LOCUS	2520 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630053H17 product:HOTTI PROTEIN homolog [Homo sapiens], full insert sequence.
ACCESSION	AK080321
VERSION	AK080321.1 GI:26348476
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
JOURNAL	High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2

TITLE
JOURNAL Normalization and subtraction of cap-trapper-selected cDNAs to
MEDLINE prepare full-length cDNA libraries for rapid discovery of new genes
PUBMED Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE 20499374
AUTHORS 11042159
 3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2520)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
 Fukuda,S., Furuno,K., Hanagaki,T., Hara,A., Hashizume,W.,
 Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
 Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
 Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Ohsato,N.,
 Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
 Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
 Takeda,Y., Tanaka,Y., Tomaru,A., Toya,T., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
 URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL: <http://genome.gsc.riken.jp/>
 URL: <http://fantom.gsc.riken.jp/>.
 Location/Qualifiers
 1. .2520
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:A630053H17"
 /db_xref="taxon:10090"
 /clone="A630053H17"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"

CDS	84..2198	/note="unnamed protein product; HOTTL PROTEIN homolog [Homo sapiens] (SPR[Q9UI99], evidence: FASTA, 85.8%ID, 93.4%length, match=821)
		putative"
		/codon_start=1
		/protein_id="BAC37878.1"
		/db_xref="GI:2634847"
		/translation="MDTSPPFYAAGQGEASPPSPVLSQSRMVRNTPYLIWT TRDVLDKRLFLSKDQMINHYRAGSFTTKVGLCLNRLNLPWFEDADSFPRCYRLG AEDDKAFIEDFLWLTARNLVLKLVLEKSQSISIQAREEAPEDTQYKQKELVT VSSDFVDEALSACQELSLTAHNDIDKPNPLSLPDDWSQFLQRYQIVHEGSLR YLEVOVQRCEDIILQOLQNVVPLDMEGRDNIWIKPKAKSRGRGIMCMRLDMLKLV DCNPMKDGKWLIVQKVERPLIFGTFEDLROWFLVTDWNLVWPHYRDSYRFTSQ PFSKLNLSVHLNCSNIOKHLSEASCHRPMLPFDNWSQRFQHLVDEADPAWSQ VIVPGRKAAVHQALQVDSQNVCRKASFELYGDFVGFQFQWLLIEINASPTMAPST AVTARACAGVQADLTWLDRLDRSDCFGAFLIYKQPAVEPQVGVIRLLVSGSTI KKPVPVGRHRTGVRSSLPHTLTOQSGESKDPSPHRSKRNKRAESLEHTEKPEP AAVASVSGKGGKAPFPSPSLKAWLPSPRVHRPQGRVLRLOHDLQVSGSKALSTGKA LMTLPKAVLMSPPHPDLKLPASMLKPKGVFELCTTWRVLSGGISGEHQRQAA PRPSAPGKLSSTFCSKTE"
ORIGIN		
Query Match	24.6%; Score 399; DB 3; Length 2520;	
Best Local Similarity	61.8%; Pred. No. 2.6e-77;	
Matches	658; Conservative 0; Mismatches 395; Indels 12; Gaps 1;	
QY	139	AGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGCAAGTGTGCCAGGCTAC 198
DB	552	AGACTAGTGACCGTCTCCTCAGACTTGTTCGATGAGGCTCTGAGTCATGCCAGGACAC 611
QY	199	CTGGGCGAGCTGAGCATGAGGACATGCACAGTCAGCAGATCCGCTGGAGGACCTCACT 258
DB	612	CTTAGCAGCATAGCCACAAAGACATGTGCAAGGACCCCAACTCCCGCTCTCACTCAGC 671
QY	259	GAGCGCGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTCATGCGCATGTTTC 318
DB	672	CCGATGATGTGTCACAGTCTCTTACGCGTACTACAAATAGTTTCATGAGGGGACAG 731
QY	319	ATCTCAATTCAGAAATTAATTTTCGAGTCCAGGCTCTGCTGAATAGATCAAGTCT 378
DB	732	CTCAGATACCTGAGGTCCTCAAGTCCAGCGCTGTGAAGACATCTCAGCAGCTTCAGAAC 791
QY	379	GTGAACCTTCAGACGACATGTAGCGGCTCCGGAACATCTGGATTATAAGCCCGGGCC 438
DB	792	GTGGTACCCAGTTAGACATGTGAGGGGGATCGGAACATCTGGATCTGGAAGCCCGAGCC 851
QY	439	AGTCCCGGGCGGACACATAGTGTGATGACCGCTGTGGAGGAGATCTCGAGCTGGCA 498
DB	852	AGTCCCGAGGCGGAGGATTAATGTGATGAACCGCTTGATGATGATGCTGAAGCTGGTG 911
QY	499	GCTGCAGACCCCTCTTTTCAGGGACAAACAGTGGGTGGTCCAGAAAGTACATCGAGACG 558
DB	912	GACTGCAACCCCATGCTCATGAAGGATGGCAAGTGGATCGTGCAGAAATACATTGAGCG 971
QY	559	CGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCACGAGCTGG 618
DB	972	CCCTCTTATCTTCGCGCACCAAGTGTGACCTGAGACAGTGGTTCCTGGTGAAGTACTG 1031
QY	619	AACCCCTGACCATCTGTTCTACAAGGAGTACTTGGCGTTCCTCAACTCAGCGCTTC 678
DB	1032	AACCCACTACCGTGTGGTTCACGAGACAGTACATTGCTTCTCCACAGGCCCTTC 1091
QY	679	TCCTGGACAAGCTGGACGCGCATCCACTGTGCAACACCGCGTCCAGAAAGTACCTG 738
DB	1092	TCCTGAAGAACCTGGACAACCTCTGTGCACTCTGTGTAACAACCTCCATCCAGACACTTG 1151
QY	739	AGAAGATGATGGGCGGCGAGCCCTCTGTCGCGGACACAAACATGTGGACGACACAGG 798
DB	1152	GAGGCTCTCTGTCAACGGGACCCGATGCTGCCCCCAGATAACATGTGGTCCAGCCAGAG 1211
QY	799	TTCCAGGATACCTGACGCGCAGGCGGTGGCGCGCTGTGGGCGGCTCATCTACCG 858

Db	1212	TTTCAGGCCCACTTTCAGGAGGTAGATGCCCAAGGCGCTGGTCCAGCGTTCATCGTGCCA 1271
QY	859	TCATGAAGAGGCCATCGCCACGCGCATGAAGTGGCCGACAGCACACGCTGGAGCCTCGC 918
Db	1272	GGCATGAAGGCTGCTGTGATCCATGCGCTGCAGACCTCCCAAGATAACGCTGCGCGG 1331
QY	919	AGAAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGGAGGAGCTTCAGGCGCTGG 978
Db	1332	AGGCCAGCTTTTGTGCTCTATGGGGGAGACTTTGTGTTTGGGGAAGACTTCCAGCGCTGG 1391
QY	979	CTGATCGAGATCAATTCAGCCCCCACCACCATGCCCGCTCCACGCGGCTCAGCGGCCAGCTG 1038
Db	1392	TTGATTTGAATCAATGCCAGCCCCCACCACCATGCCCGCTTCACAGCGCTGTCACTGCGCGCTC 1451
QY	1039	TGTGCACAGGTGCAGAGGAGACAC-----CATCAAGTGGCGCTGGACCGGAGC 1086
Db	1452	TGTGCCGGTGTGCAGGCAGATACCTTCGCTGCTGTCATCGACCGGCGCTGGACCGGAGC 1511
QY	1087	TGTGACATCGGCACTTCAGAGCTTCCTGTGGAGGACCGGTGCTTGGAGCGCGCCCATTC 1146
Db	1512	TGTGATACGGAGGCTTTGAGCTCATCTATAGCACCTGCTGTGGAGTGCCTCAGTAC 1571
QY	1147	AGCGGTCGCGACCTTCGCTGGCGGCGCTCAGTGTGAGGAGGCC 1191
Db	1572	GTGGGTATCGGCTCTAGTGGAGGCTCTACCATCAAGAAGCCC 1616
RESULT 11		
BM808637		
LOCUS	BM808637	
DEFINITION	AGENCOURT_6617786 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734232	
ACCESSION	BM808637	
VERSION	BM808637.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	1 (bases 1 to 1023)	
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: Invitrogen cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12739 row: d column: 09 High quality sequence start: 6 High quality sequence stop: 664. Location/Qualifiers 1..1023 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5734232" /tissue_type="hippocampus" /lab_host="DH10B" /clone_lib="NIH_MGC_124" /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcorV (destroyed); Site 2: NotI; RNA source: male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."	
FEATURES		
source		
ORIGIN		

Query Match		23.3%;	Score 378.4;	DB 4;	Length 1023;
Best Local Similarity		69.5%;	Pred. No. 8.1e-73;		
Matches		514;	Conservative	0;	Mismatches 226; Indels 0; Gaps 0;
QY	345	GCAGTCCAGGCTCTGCTGAATAGATACAGCTGTGTGAACCTCAGACGGACATTGACGG	404		
DB	49	CGCGTGTGAGGACATCTCTGCAGCAGCTGCAGGCGGTGTACCCAGATAGATGAGGAG	108		
QY	405	GCTCCGGAACATCTGATATATAAGCCCGCCGCAAGTCCCGGGCCGAGACATAGTGTG	464		
DB	109	GGATGCAACATCTGGATCGTGAAGCCAGAGACCAAGTCCCGGACGAGGATCATGTG	168		
QY	465	CATGACCGTGTGGAGGAGATCTGTGAGCTGGCAGCTGCAGACCAACCTCTTTTCCAGGA	524		
DB	169	CATGACCACTGGAGGAGATCTGAAGCTGTGAAGGCAACCCGCTGTGATGAAGA	228		
QY	525	CAACAAGTGGTGGTCCAGAAGTACATGAGAGCGGCTGTCTATCTGTGACACCAAGTT	584		
DB	229	CGGCAAGTGGTGGTGCAGAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAAGTT	288		
QY	585	GCACATCAGACAGTGTCTCTCGTCACGACCTGGAACCCCTGACCATCTGTTCTACAA	644		
DB	289	TGACCTCAGACAGTGTCTCTGTTAACTGACTGGAAACCACTTACCGTGTGTTTACCG	348		
QY	645	GGAGAGTTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACGCGCAT	704		
DB	349	GCACAGCTATATCCGCTTTTCCAGCAGCGCTTCTCCCTGGAAGAACCTGGACAACCTCAGT	408		
QY	705	CCACTGTGCACAACGCGCTGCAGAAGTACCTGAAGATGATGTGGCGGACGAGCCCT	764		
DB	409	GCACCTGTGCAACAACTCCATCCAGAAGCACCTGGAGAACCTCATGCCATCCACT	468		
QY	765	GCTGCCCGCACACAATGTGACACAGCACCAAGTCTCAGGAGTACCTGACGCGCAGGG	824		
DB	469	GCTTCGCGCAGACAATGTGTCTAGCCAGAGGTTCCAGGCCCACTTGCAGAGATGGG	528		
QY	825	CCGTGGCGCGGTGTGGGCGAGCTCATCTACCCGCTCCATGAAGAAGGCCATCGCCACGC	884		
DB	529	TGCCCCAAATGTTGGTCACCATCATCTGTGCTGGCATGAAGGATGCTGTGATCCACGC	588		
QY	885	CATGAAGTGGCCAGGACCACTGTGAGCTGTGATCGAGATCAATTCAGCCCCAC	944		
DB	599	ACTTCAGACCTCCCAAGACACCTGCAGTGTGCGAAGCCAGCTTTGAGCTCTATGGCGC	648		
QY	945	TGACTTCGTCTGTGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCCCAC	1004		
DB	649	TGACTTCGTCTGTGGAGGACTTCAGGCCCTGGCTGATCGAGATCAACGCCAGCCCCAC	708		
QY	1005	CATGACCCGCTCCAGCCGCTCACGCCCGAGCTGTGTGCACAGGTGCGAGGAGCACCAT	1064		
DB	709	GATGCAACCTCCACAGCAGTCACTGCCCGGCTCTGTGTGCGCGTGCAAGCTGACACCT	768		
QY	1065	CAAGTGGCGGTGGACCGCA	1084		
DB	769	GCGCGTGGTCAATTGACCGGA	788		

RESULT 12
BM808516
LOCUS
DEFINITION
AGENCY 6617783 NIH_MGC_124 Homo sapiens cdna clone IMAGE:5734160
5', mRNA sequence.
ACCESSION
BM808516
VERSION
BM808516.1 GI:19125339
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1022)
AUTHORS
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12739 row: a column: 09
High quality sequence start: 78
High quality sequence stop: 658.
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734160"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV (destroyed); Site_2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

ORIGIN		Query Match	22.2%;	Score 360;	DB 4;	Length 1022;
		Best Local Similarity	68.8%;	Pred. No. 9.4e-69;		
		Matches	495;	Conservative	0;	Mismatches 225; Indels 0; Gaps 0;
QY	345	GCAGTCCAGGCTCTGTGAATAGAAATCACGTCTGTGAACCTCAGACGGACATTGACGG	404			
DB	40	GCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCTGTATCCAGATAGACATGAAAGG	99			
QY	405	GCTCCGGAACATCTGATATATAAGCCCGCCGCAAGTCCCGGGCCGAGACATAGTGTG	464			
DB	100	GGATCGCAACATCTGTGATCGTGAAGCAGGAGCCAAAGTCCCGCGCAGGAGCATCATGTG	159			
QY	465	CATGACCGCTGTGAGGAGATCCTCGAGCTGCAGCTGCAGACCAACCTCTTTTCCAGGA	524			
DB	160	CATGACCACTGTGAGGAGATGCTGAAGCTGGTGAACCGCAACCCGCTGGTATGAGGA	219			
QY	525	CAACAAGTGGTGTGCAGAAAGTACATCGAGACGCGCTGTCTATCTGTGACACCAAGTT	584			
DB	220	CGGCAAGTGGTGTGTCAGAAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAAGTT	279			
QY	585	CGACATCAGACAGTGTTCCTCGTCACGAGCTGGAAACCCCTGACCATCTGGTCTACAA	644			
DB	280	TGACCTCTTACAGTGGTTCCTGGTAACTGACTGGAAACCCACTTACCGTGTGTTTACCG	339			
QY	645	GGAGAGTTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTCGACAAGCTGGACAGGCCAT	704			
DB	340	CGACAGCTATATCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAACCCTGGACAACCTCAGT	399			
QY	705	CCACTGTGCAACAAACGCGCTGCAGAAAGTACCTGAAGAATGATGTGGGCGCAGGCCCT	764			
DB	400	GCACCTGTGCAACAACTCCATCCAGAAAGCACTCGAGAACTCATGCCATCGGCATCCACT	459			
QY	765	GCTGCCCGCACACAATGTGGAACAGCACGAGGTTCCAGGAGTACCTGCGCGCAGGG	824			
DB	460	GCTTCCCGCAGACAACATGTGTTCTAGCCAGAGGTTCCAGGCGCCACCTGCGAGAGATGGG	519			
QY	825	CCGTGGCGCGGTGTGGGCGAGCGTTCATCTACCCGCTCCATGAAGAAGGCCATCGCCACGC	884			
DB	520	TGCCCCAAATGCTTGGTCCACCATCATCTGCTGCTGCGCATGAAGGATGCTGTGATCCAGC	579			
QY	885	CATGAAGTGGCCCGGACGAGCCCTCGCAAGAACAGCTTTGAGCTCTACGGGC	944			

Db 580 ACTTCAGACCTCCAGGACACCGTGCAGTGTGGAGGCCAGCTTTCAGCTCTATGGCG 639
 QY 945 TGACTTCGTCTTGGAGGAGGACTTCAGGCTTGGCTGTGATCGAGATCAATTCAGCCCCAC 1004
 Db 640 TGACTTCGTCTTGGAGGAGGACTTCAGGCTTGGCTGTGATCGAGATCAATTCAGCCCCAC 699
 QY 1005 CATGACCCCGTCCACCGCGTTCAGCGCCAGCTGTGACAGGTGTCAGGAGGACACCAT 1064
 Db 700 GATGGCACCCCTCACAGCAGTCACTGGCCGCGCTCTGTGCTGGCGTGAAGCTGACACCC 759

RESULT 13

BM717061

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BM717061 693 bp mRNA linear EST 28-FEB-2002
 UI-E-EJO-ahk-e-05-0-UI-r2 UI-E-EJO Homo sapiens cDNA clone
 UI-E-EJO-ahk-e-05-0-UI 5', mRNA sequence.

BM717061
 BM717061.1 GI:19030319
 EST.
 Homo sapiens (human)
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 693)

Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery

Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548

Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers

FEATURES

source

1..693
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-EJO-ahk-e-05-0-UI"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-EJO"
 /notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-EJO is a subcloned cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT7T3-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes,
 AGAATCAAGA; lens, CCATAGGCA; eye anterior segment,
 AATGCCGAT; optic nerve, CCATTAAGT; retina, CCGCG; Retina
 Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
 library was created for the program, Gene Discovery in the

Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 19.8%; Score 321; DB 4; Length 693;
 Best Local Similarity 69.1%; Pred. No. 3.7e-60;
 Matches 473; Conservative 0; Mismatches 200; Indels 12; Gaps 2;

QY 451 CGAGACATAGTGTGATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAC 510
 Db 1 CGAGGACATGATGTGCATGGACACACCTGGAGGAGATGCTGAAGCTGGT-GAACGGCAGCCC 59

QY 511 CCTCTTTCCAGGACAAACAAAGTGGGTGGTCCAGAAGTACATCGAGAGCCGCTGCTCATC 570
 Db 60 GTGGTATGAGGACCGCAAGTGGTGGTGCAGAGTATATTGAGCGGCCCTCTCTCATC 119

QY 571 TGTGACACCAAGTTCACATCAGACAGTGTTCCTCGTCACGAGTGGAAACCCCTGAGCC 630
 Db 120 TTTGGCACCAGTTTGACCTCAGACAGTGTTCCTGTTAACTGACTGGAACCCACTTACC 179

QY 631 ATCTGGTTCTACAGGAGAGTTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTGGACAAG 690
 Db 180 GTGTGGTTCTACCGGACAGCTATATCCGCTTTCCACGAGAGGTTTCCAGGCCAC 239

QY 691 CTGGACAGCGCCATCCACCTGTGCAACACGCGTCCAGAAGTACCTGGAAGATGATGTG 750
 Db 240 CTGGACACTCAGTGCACCTGTGCAACAACTCATCCAGAGCACTTGGAGAACTCATGC 299

QY 751 GGCCTGAGCCCTCTGCTGCCCGCACCAACATGTGGACAGCACAGGTTTCCAGGAGTAC 810
 Db 300 CATCGGCATCCACTGCTTCGGCCAGACAAACATGTGTGTAGCCAGAGGTTTCCAGGCCAC 359

QY 811 CTGCAGCGCCAGGCGCGTGTGGCGCGCTGTGGGAGAGCTATCTACCGTCCATGAAGAAG 870
 Db 360 CTGCAGGAGATGGGTGCCCAAAATGCTGTGTCACCATCATCGTGTGCTGGCATGAAGAT 419

QY 871 GCCATCGCCAGCCCATGAAGTGGCGCAGGACCAAGTGGAGCTCGCAGAGCAAGCTTT 930
 Db 420 GCTGTGATCCAGCATTCAGACCTCCAGGACACCGTGCAGTGTGGAGGCGAGCTTT 479

QY 931 GAGCTCTACGGGCTGACTTCGTCTTGGAGGAGACTTCAGGCCCTGGCTGTGATCGAGATC 990
 Db 480 GAGCTCTATGGCGTGCATTCGTGTTTCGGGGAGGACTTCCAGCCCTGGCTGTGATGATC 539

QY 991 AATTCAGCCCAACCATGACCCGCTCCAGCCGCTCAGCCGCTGAGTGTGTCACAGGTG 1050
 Db 540 AAGCCAGCCCAACCATGACCCCTCCAGCAGTCACTGCCCGGCTCTGTGTGGCGTG 599

QY 1051 CAGGAGGACAC-----CATCAAGTGGCGCTGGAGCGCAGCTGTGACATCGGCA 1099
 Db 600 CAAGCTGAACCTGCGCGTGTGATGACCGGATGTCGACCCCAACTGTGACACAGGAG 659

QY 1100 ACTTCGAGCTCTGTGGAGGCGAGCC 1124
 Db 660 CCTTTGAGCTCATCTATAAGCAGGC 684

RESULT 14
 CK653583

LOCUS

DEFINITION

AGENCY: 17673480 NIH_MGC_237 Rattus norvegicus cDNA clone

IMAGE: 7113394 5', mRNA sequence.

CK653583

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 861)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CK653583 861 bp mRNA linear EST 29-JAN-2004

AGENCY: 17673480 NIH_MGC_237 Rattus norvegicus cDNA clone

IMAGE: 7113394 5', mRNA sequence.

CK653583

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 861)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-remail.nih.gov Tissue Procurement: Howard Jacobs cDNA Library Preparation: Express Genomics cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM14985 row: e column: 08 High quality sequence stop: 304. Location/Qualifiers		491 ATTATCTTCGATCTCTATGGAGCTGACTTTAATCTGGGCGTGACTTATAGTCTCTGTGT 550 981 GATCGAGATCAATTCCAGCCCCACCATGCACCGCTCCAGCCGCGTCACGGCC 1032 551 CTTCAAGATCAACTTCTACTTGCCTTGCACCTCTACTTTTGTAAACACC 602	
	Location/Qualifiers 1. 861 /organism="Rattus norvegicus" /mol_type="mRNA" /db_xref="taxon:10116" /clone="IMAGE:7113394" /tissue_type="testis, pooled" /lab_host="DH10B Tona" /clone_lib="NIH MGC 237" /note="Organ: testis; Vector: pExpress-1; Site.1: EcoRV; Site.2: NotI; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTAGTCTAGATCGGAGCGCGCCCT(7)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb resulted in an average insert size of 2.4 kb. This primary library is not normalized (normalized primary library is NIH MGC 238) and was constructed by Express Genomics (Frederick, MD)"		CD515431 896 bp mRNA linear EST 06-JUN-2003 AGENCOURT_14364765 NIH_MGC_181 Homo sapiens cDNA clone IMAGE:30407505 5', mRNA sequence. CD515431 EST. CD515431.1 GI:31447149 SOURCE Homo sapiens (human) ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE 1 (bases 1 to 896) NIH-MGC http://mgc.nci.nih.gov/ AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC) TITLE Unpublished (1999) JOURNAL Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: NDAM488 row: p column: 10 High quality sequence stop: 678. Location/Qualifiers	
FEATURES	source		1. 896 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:30407505" /tissue_type="White Matter" /dev_stage="Unknown" /lab_host="DH10B-Ton A (T1 and T5 phage resistances) " /clone_lib="NIH_MGC_181" /note="Vector: pCMV-Sport6.1; Site.1: NotI; Site.2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.42 kb. Library was constructed by Invitrogen). Note: this is a NIH_MGC Library."	
	Query Match 19.0%; Score 307.6; DB 6; Length 896; Best Local Similarity 70.0%; Pred. No. 3.5e-57; Matches 434; Conservative 0; Mismatches 174; Indels 12; Gaps 1;		520 AGGAGCAACAAGTGGTGTCCAGAAGTACATCGAGACCGCGCTGCTCATCTGTGACACC 579 25 AAGGACGGCAAGTGGTGTGCGAAGTATATTGAGCGCGCCCTCTCTCATCTTTGGCACC 84 580 AAGTTTCGACATCAGACAGTGGTTCTCGTCACGGAAGTGGAAACCCCTGACCATCTGGTTTC 639 85 AAGTTTGACCTCAGACAGTGGTTCTCTGTTAACTGAGTGAACCCACTTACCGTGTGGTTTC 144 640 TACAAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGC 699 145 TACCGCGACAGCTATATCGGCTTTTCCAGCAGCGCTTCTCCCTGAAGAACCTGGACAAC 204 700 GCCATCCACCTGTGCAACACCGCCCTCCAGAGTACCTGAAGAATCATGTGGCGCCGACG 759	
ORIGIN	source		Query Match 19.5%; Score 316; DB 7; Length 861; Best Local Similarity 74.2%; Pred. No. 4.8e-59; Matches 439; Conservative 0; Mismatches 150; Indels 3; Gaps 3;	
	443 CCGGGGCGGACATAGTGTGATGGACCGGTGGAGGAGATCTTGGAGCTGGGAGCTG 502 12 CCGGATTTGGAGATATTGTGATGGACCGGTGGAGAACATCTGGATCTGGTGGCC- 70 503 CAGACACCCCTCTTTCCAGGGAACAAGTGGGTGTGCAGAAATACATCGAGACCGCCG 562 71 CAGACAGGACAGACCAGGAGGACACAATGGGTGTCATAAGTACATCGAGACCCCGA 130 563 TGCTCATCTGTGACACCAAGTTGACATCAGACAGTGGTTCTCGTCACGGAAGTGAACC 622 131 TGCTCATCTATGACACCAAGTTGACATCAGACAGTGGTTCTTGTTCACAGACTGGAATC 190 623 CCCTGACCATCTGTTCTACAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCC 682 191 CCTTAACCATCTGGTTCTACAAGAGAGTACTCTCGGTTCTCCAGCGCGCTTCTCCC 250 683 TGGACAAGCTGGACAGCGCATCTCCACTTGCAACAACGCCGCTCCAGAGTACCTGAAGA 742 251 TGGACAAACTGACAGCGGCTTCCACTGTGTAACTCCGTCAGAGCGGTGGCATAA 310 743 ATGATGTGGGCGGACGCCCTTGTGTCGCCGCAACAACATGTGGACCAAGCAAGGTTCC 802 311 ATGACAAAAAAGCAAGTCCACTGCTACCTTGGCCATAAACAATGTGGACCATCACCGCTTCC 370 803 AGGAGTACTCTGAGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGTCATCTACCGGTCCA 862 371 ATGAGTACTCTGATTAATAGGGCCCGAGGATGCACATGGGGTAGTATCATCTACCCGCTTA 430 863 TGAAGAAGGCCCATCGC-CCACGCGCATGAAGGTGGGCCAGGACCAAGTGGAGCGCTGGCAAG 921 431 TGAAGAAGAGCTGTCACTCAACGTCATGATGGTGGGCCCAAGATCTTGTGAAGCCCGTAG 490 922 AACAGCTTTGAGCTCTACGGGGCTGACTTCCTGCTTTGGGAGGAGCTT-CAGGCGCTGGCT 980		520 AGGAGCAACAAGTGGTGTCCAGAAGTACATCGAGACCGCGCTGCTCATCTGTGACACC 579 25 AAGGACGGCAAGTGGTGTGCGAAGTATATTGAGCGCGCCCTCTCTCATCTTTGGCACC 84 580 AAGTTTCGACATCAGACAGTGGTTCTCGTCACGGAAGTGGAAACCCCTGACCATCTGGTTTC 639 85 AAGTTTGACCTCAGACAGTGGTTCTCTGTTAACTGAGTGAACCCACTTACCGTGTGGTTTC 144 640 TACAAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGC 699 145 TACCGCGACAGCTATATCGGCTTTTCCAGCAGCGCTTCTCCCTGAAGAACCTGGACAAC 204 700 GCCATCCACCTGTGCAACACCGCCCTCCAGAGTACCTGAAGAATCATGTGGCGCCGACG 759	

```
Db 205 TCAGTGACCTGTGCAACAACTCCATCCAGAACACCTGGAGAACTCATGCCATCGGCAT 264
Qy 760 CCCTCTGCTCCCGCAGACACATGTGGACACAGACACAGGTTCCAGGAGTACCTGCGAGCGC 819
Db 265 CCACTGCTTCCGCCAGACAAACATGTGGTCTAGCCAGAGGTTCCAGGCGCCACCTGCGAGGAG 324
Qy 820 CAGGGCCGTGGCGCGTGTGGGCGAGCGTCTATCCCGCTCCATGAAGAGGCCATCGCC 879
Db 325 ATGGTGCCCAAAATGCTTGGTCCACCATCATGTGCTGGCATGAAGGATGCTGTGATC 384
Qy 880 CACGCCATGAAGGTGGCCAGGACCAACGTGGAGCTCGCAAGAACAGCTTTGAGCTCTAC 939
Db 385 CACGCACTTCAGACCTCCAGGACACCGTGCATGTGCGAAGGCCAGCTTTGAGCTCTAT 444
Qy 940 GGGGCTGACTTCGCTCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTCAGC 999
Db 445 GGGGCTGACTTCGCTCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAACGCCAGC 504
Qy 1000 CCCACCATGCACCCGTCCACGCCGTCCAGGCCAGCTGTGTCACAGGTGCAGGAGGAC 1059
Db 505 CCCACGATGGCACCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGCAAGCTGAC 564
Qy 1060 ACCATCAAGGTGGCC-----GTGGACCGCAGCTGTGACATCGGCAACTTCGAG 1107
Db 565 ACCCTGCGCGTGTGTCATTGACCGGAGGCTGGACCGCAACTGTGACACAGGAGCCTTGAG 624
Qy 1108 CTCCTGTGGAGGAGCGCGT 1127
Db 625 CTCATCTATAAGCAGGCCGT 644
```

Search completed: September 25, 2005, 01:56:01
Job time : 6239.23 secs

This page blank (uspto)

no.	score	match length	pos	description
1	1623	100.0	1838	ADJ93357
2	1619.8	99.8	1838	ADJ93364
3	1619.8	92.8	1859	ADJ93362
4	1475.8	90.9	1845	ADJ93363
5	1475.8	90.8	1939	ADJ93361
6	726	44.7	726	ADJ93359
7	501	30.9	101270	ADQ17814
8	432.8	26.7	3280	AAS99894
9	431.6	26.6	2326	AAR16735
10	427	26.3	490	ACH91699
11	421.4	26.0	2538	FADS73168
12	421.4	26.0	2553	ACN37881
13	421.4	26.0	2881	AACT7214
14	421.4	26.0	3001	AAI58606
15	421.4	26.0	3001	ADQ98824
16	421.4	26.0	3001	ADB49584
17	419.8	25.9	3294	ABX34502
18	418	25.8	2848	ADM03081
19	413	25.4	2611	AAF21812
20	347.4	21.4	2412	ADH45468
				Adj93357 Human BGS
				Adj93364 Human tub
				Adj93362 Human BGS
				Adj93363 Human BGS
				Adj93361 Human BGS
				Adj93359 Human BGS
				Adj93359 Human BGS
				Adj93359 Human sof
				Adq17814 Human nuc
				Aas99894 Polynucle
				Aar16735 Human cDN
				Ach91699 Human gen
				Fads73168 Human kid
				Acn37881 Tumour-as
				Aac77214 Human ORF
				Aai58606 Human pol
				Adg98824 DNA encod
				Adb49584 Novel hum
				Abx34502 Human mdt
				Adm03081 Human cDN
				Aaf21812 Human bre
				Adh45468 Human mol

XX
PT
Feder .IN
Wils Nelson TC.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.

XX WPI; 2004-099381/10.
DR P-PSDB; ADJ93358.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
disorders.
XX
PS Claim 1; SEQ ID NO 1; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.
XX
SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Query Match 100.0%; Score 1623; DB 12; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGGCCACAGAGCTGCAGAGGACGACGAGA 60
DB |||||
DB 153 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGGCCACAGAGCTGCAGAGGACGACGAGA 212
QY 61 AGCAAGCCAGGACCCAGAGGGAGGAGGCGGGAGCAGCGACTGAGCAGAGGCAAGAT 120
DB |||||
DB 213 AGCAAGCCAGGACCCAGAGGGAGGAGGCGGGAGCAGCGACTGAGCAGAGGCAAGAT 272
QY 121 GCTGAAATGCTCAGGCAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGACATCGGGTGC 180
DB |||||
DB 273 GCTGAAATGCTCAGGCAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGACATCGGGTGC 332
QY 181 AAGGTGTGCGAGGCTTACCTGGGGAGCTGGAGCATGAGGACATGACACGTCAGCAGAT 240
DB |||||
DB 333 AAGGTGTGCGAGGCTTACCTGGGGAGCTGGAGCATGAGGACATGACACGTCAGCAGAT 392
QY 241 GCCGTGGAGGACCTCACTAGGCGCGAGTGGGAGGACCTGACCCAGCAGTACTACCTCCCTC 300
DB |||||
DB 393 GCCGTGGAGGACCTCACTAGGCGCGAGTGGGAGGACCTGACCCAGCAGTACTACCTCCCTC 452
QY 301 GTTCATGGCGATGCTTTCATCTCAATTCAGAAATTAATTTTCGAGTGCCAGGCTCTG 360
DB |||||
DB 453 GTTCATGGCGATGCTTTCATCTCAATTCAGAAATTAATTTTCGAGTGCCAGGCTCTG 512
QY 361 CTGAATAGAAATCAGCTCTGTGAACCTCAGACGGACATTCAGCGGGCTCCGGAAACATCTGG 420
DB |||||
DB 513 CTGAATAGAAATCAGCTCTGTGAACCTCAGACGGACATTCAGCGGGCTCCGGAAACATCTGG 572
QY 421 ATTATAAAGCCCGCGCAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCCGCTGTGGAG 480

DB 573 ATTATAAAGCCCGCGCCAAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCGCTGTGGAG 632
QY |||||
QY 481 GAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACCAACAGTGGGTGGTC 540
DB |||||
DB 633 GAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACCAACAGTGGGTGGTC 692
QY |||||
QY 541 CAGAAGTACATCAGAGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
DB |||||
DB 693 CAGAAGTACATCAGAGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 752
QY |||||
QY 601 TTCCTCGTCAAGGATGGAACCCCTGACCATCTGTTTCTTCAAGGAGAGTACTTTGGCG 660
DB |||||
DB 753 TTCCTCGTCAAGGATGGAACCCCTGACCATCTGTTTCTTCAAGGAGAGTACTTTGGCG 812
QY |||||
QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 720
DB |||||
DB 813 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 872
QY |||||
QY 721 GCCGTCCAGAAGTACCTGAAGAAATGATGTGGGCGCGAGCCCTCTGTGCCCGCACACAAC 780
DB |||||
DB 873 GCCGTCCAGAAGTACCTGAAGAAATGATGTGGGCGCGAGCCCTCTGTGCCCGCACACAAC 932
QY |||||
QY 781 ATGTGACCCAGCACCAAGTTCAGAGATACCTTGACGCGCCAGGCGCGTGGCCGCTGTGG 840
DB |||||
DB 933 ATGTGACCCAGCACCAAGTTCAGAGATACCTTGACGCGCCAGGCGCGTGGCCGCTGTGG 992
QY |||||
QY 841 GCGAGGTCTATCTACCCGTCATGAGAGGCCATCGCCACGCCCATGAAGTGGGCCACAG 900
DB |||||
DB 993 GCGAGGTCTATCTACCCGTCATGAGAGGCCATCGCCACGCCCATGAAGTGGGCCACAG 1052
QY |||||
QY 901 GACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG 960
DB |||||
DB 1053 GACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCCTTGGG 1112
QY |||||
QY 961 AGGAGCTTCAGGCGCTGGCTGATTCAGATCAATTCAGGCCCCACCATGACCCGCTCCAG 1020
DB |||||
DB 1113 AGGAGCTTCAGGCGCTGGCTGATTCAGATCAATTCAGGCCCCACCATGACCCGCTCCAG 1172
QY |||||
QY 1021 CCGGTTCAGGCCCCAGCTGTGCACAGGTGCAGGAGGACACCATCAAGTGGCCGCTGGAC 1080
DB |||||
DB 1173 CCGGTTCAGGCCCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGTGGCCGCTGGAC 1232
QY |||||
QY 1081 CCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGTGAGGAGCAGCGGTGTTGAGCGGCC 1140
DB |||||
DB 1233 CCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGTGAGGAGCAGCGGTGTTGAGCGGCC 1292
QY |||||
QY 1141 CCATTTCAGGCGGTTCGACCTCTGCGTGGCGGGCTCAGTGTGAGGAGAGCCAGGAGGACAG 1200
DB |||||
DB 1293 CCATTTCAGGCGGTTCGACCTCTGCGTGGCGGGCTCAGTGTGAGGAGAGCCAGGAGGACAG 1352
QY |||||
QY 1201 GTGCTGCCGCTGTGCAACTCAAGGCTTCGAGCTCTGTGTGAGGAGCAGCGGTGTTGAGG 1260
DB |||||
DB 1353 GTGCTGCCGCTGTGCAACTCAAGGCTTCGAGCTCTGTGTGAGGAGCAGCGGTGTTGAGG 1412
QY |||||
QY 1261 GCACGGGGGGCCCCCTCGGCCATGCGAGACCTGTGCCAGGAGACCCCATCACAGCTCTCCAG 1320
DB |||||
DB 1413 GCACGGGGGGCCCCCTCGGCCATGCGAGACCTGTGCCAGGAGACCCCATCACAGCTCTCCAG 1472
QY |||||
QY 1321 CGGGACTTTGGGACTGAAGAGAGAGGGGCTCCCTCGCTGGCCCTTGTGTGACACCTTTAAGG 1380
DB |||||
DB 1473 CGGGACTTTGGGACTGAAGAGAGAGGGGCTCCCTCGCTGGCCCTTGTGTGACACCTTTAAGG 1532
QY |||||
QY 1381 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAGAGCTCTCGGAGAGGTG 1440
DB |||||
DB 1533 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAGAGCTCTCGGAGAGGTG 1592
QY |||||
QY 1441 GAGCTCCCGGCCCTGCGCCCTGTGCGCACGCTGGAGAGTTCAGGCGGCCCAACACACCGGTGTCCC 1500
DB |||||
DB 1593 GAGCTCCCGGCCCTGCGCCCTGTGCGCACGCTGGAGAGTTCAGGCGGCCCAACACACCGGTGTCCC 1652
QY |||||
QY 1501 GTAGCCGAGCCCGCCCAAGAGCTGGGATCCAAACAGAGCTTAATTCGCGCACCCGCTGTGAGCCT 1560
DB |||||

Db	1553	GTAGCCACCGCCGCAAAAGCTGGGATCCAAACACAGCTAAATGGCGCACCCGCTGGAGCCT	1711
Qy	1561	GTGCTGGGGGCTGAAGACAGCAGCAGAGGGCGCGCTGCGTCCGCGCCCGGAGGAAAGGT	1620
Db	1713	GTGCTGGGGGCTGAAGACAGCAGCAGAGGGCGCGCTGCGTCCGCGCCCGGAGGAAAGGT	1772
Qy	1621	TCA 1623	
Db	1773	TCA 1775	
RESULT 2			
ADJ93364			
ID	ADU93364	standard; DNA; 3554 BP.	
XX	ADJ93364;		
XX			
DT	06-MAY-2004	(first entry)	
XX			
DE	Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.		
XX			
KW	testis-specific tubulin tyrosine-ligase-like polypeptide;		
KW	BGS-42 polypeptide; cytotatic; respiratory-Gen; gastrointestinal-Gen;		
KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;		
KW	osteoprotic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;		
KW	anti-HIV; antibacterial; immunosuppressive; antiseborrheic;		
KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;		
KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;		
KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;		
KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;		
KW	brain cancer; liver cancer; proliferative condition; testis; lung;		
KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;		
KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;		
KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;		
KW	sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.		
XX			
OS	Homo sapiens.		
OS	Synthetic.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	1695..3320	
FT		/*cag= a	
FT		/product= "Human tubulin tyrosine ligase protein"	
XX			
PN	WO2004005487-A2.		
XX			
PD	15-JAN-2004.		
XX			
PF	09-JUL-2003; 2003WO-US021605.		
XX			
PR	09-JUL-2002; 2002US-0394725P.		
XX			
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.		
XX			
PI	Feder JN, Wu S, Nelson TC;		
XX			
XX	WPI; 2004-099381/10.		
DR	P-PSDB; ADJ93365.		
XX			
PT	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,		
PT	useful for preventing, treating or ameliorating a medical condition, e.g.		
PT	aberrant cellular proliferation, reproductive disorders or testicular		
PT	disorders.		
XX			
PS	Example 4; SEQ ID NO 12; 343pp; English.		
XX			
CC	This invention relates to a novel testis-specific tubulin tyrosine-ligase		
CC	-like polypeptide, designated the BGS-42 polypeptide. The invention may		
CC	be useful for the development of compounds with a cytotatic, respiratory		
CC	-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,		
CC	antiinflammatory, anabolic, hypertensive, osteoprotic, nootropic,		
CC	antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,		
CC	immunosuppressive, antiseborrheic or dermatological activity acting as		

	tyrosine ligase modulators. In addition, the disclosed sequences may be useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of the tubulin tyrosine ligase protein consensus gene sequence which was used in the exemplification of the invention.
CC	
XX	
SQ	Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;
	Query Match 99.8%; Score 1619.8; DB 12; Length 3554;
	Best Local Similarity 99.9%; Pred. No. 0;
	Matches 1621; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ATGGCATTCAGCATCTCAAGTGGGTGTTGAGCCACAGAGCTGCAGCAGGACGACGACAGA 60
DB	1695 ATGGCATTCAGCATCTCAAGTGGGTGTTGAGCCACAGAGCTGCAGCAGGACGACGACA 1754
QY	61 AGCAAGCCCAGGGACACAGAGGGAGGAGCGCGGAGCAGCGACTGTAGCAGCAGCGCAAGAT 120
DB	1755 AGCAAGCCCAGGGACACAGAGGGAGGAGCGCGGAGCAGCGACTGTAGCAGCAGCGCAAGAT 1814
QY	121 GCTGAATAATGCTGAGGCAAAGCTCAGGGGCTCCCGGGGAGGCTTGAGCATCGCGTGC 180
DB	1815 GCTGAATAATGCTGAGGCAAAGCTCAGGGGCTCCCGGGGAGGCTTGTTGGACATCGCGTGC 1874
QY	181 AAGGTGTGCCAGSGCTTACCCTGGSGCAGCTGGAGCATGAGGACATCGACAGCTCAGCAGAT 240
DB	1875 AAGGTGTGCCAGSGCTTACCCTGGSGCAGCTGGAGCATGAGGACATCGACAGCTCAGCAGAT 1934
QY	241 GCCTGTGGAGAACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTAGTACTACTCCCTC 300
DB	1935 GCCTGTGGAGAACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTAGTACTACTCCCTC 1994
QY	301 GTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTCTG 360
DB	1995 GTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTCTG 2054
QY	361 CTGAATAGAAATCACGTCGTGTGAACCTCTCAGCGGACATTTGACGGGCTCCGGAAACATCTGG 420
DB	2055 CTGAATAGAAATCACGTCGTGTGAACCTCTCAGCGGACATTTGACGGGCTCCGGAAACATCTGG 2114
QY	421 ATTATAAAGCCCGCGCCCAAGTCCCGGGGCGGAGACATAGTGTGSCATGGACCGTGTGGAG 480
DB	2115 ATTATAAAGCCCGCGCCCAAGTCCCGGGGCGGAGACATAGTGTGSCATGGACCGTGTGGAG 2174
QY	481 GAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTTTTCAGGGGACAAACAGTGGGTGGTTC 540
DB	2175 GAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTTTTCAGGGGACAAACAGTGGGTGGTTC 2234
QY	541 CAGAAGTACATCGAGACGGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
DB	2235 CAGAAGTACATCGAGACGGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 2294
QY	601 TTCTCTCGTCA CGGACTGGAAACCCCTTGACCATCTGGTCTTACAAAGGAGAGTACTTATCGCG 660
DB	2295 TTCTCTCGTCA CGGACTGGAAACCCCTTGACCATCTGGTCTTACAAAGGAGAGTACTTATCGCG 2354
QY	661 TTCTCAACTCAGCGCTTCTCCCTTGGACAAGCTGGACAGCGCCATCCACCTGTGCACAAC 720

Db 2355 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCCATCCACTGTGCAACAAC 2414
Qy 721 GCGCTCAGAAAGTACCTGAAGAATGATGTGGCGCGCAGCCGCCCTGTGTCGCCGACACAAC 780
Db 2415 GCGCTCAGAAAGTACCTGAAGAATGATGTGGCGCGCAGCCGCCCTGTGTCGCCGACACAAC 2474
Qy 781 ATGTGACCAAGCAACAGGTTCCAGAGTACTCTGAGGCGCCAGGCGCGTGGCGCGGTGG 840
Db 2475 ATGTGACCAAGCAACAGGTTCCAGAGTACTCTGAGGCGCCAGGCGCGTGGCGCGGTGG 2534
Qy 841 GGCAGGCTATCTACCGCTCATGAAGAGGCCATCGCCCAACGCCATGAAGTGGGCCAG 900
Db 2535 GGCAGGCTATCTACCGCTCATGAAGAGGCCATCGCCCAACGCCATGAAGTGGGCCAG 2594
Qy 901 GACACGTGAGGCTCGCAAGAACAGCTTTGAGCTCTACGGGCTGACTTCGTCCTTGGG 960
Db 2595 GACACGTGAGGCTCGCAAGAACAGCTTTGAGCTCTACGGGCTGACTTCGTCCTTGGG 2654
Qy 961 AGGGAATTCAAGGCCCTGGCTGATCGAGATCAATTCAGGCCCAACCATGCAACCGTCCAG 1020
Db 2655 AGGGAATTCAAGGCCCTGGCTGATCGAGATCAATTCAGGCCCAACCATGCAACCGTCCAG 2714
Qy 1021 CCGGTACGCCCCAGCTGTGTGCACAGGTGCAGAGGACACATCAAGTGGCCGTGGAC 1080
Db 2715 CCGGTACGCCCCAGCTGTGTGCACAGGTGCAGAGGACACATCAAGTGGCCGTGGAC 2774
Qy 1081 CGGAGCTGTGACATCGCAACTTCGAGCTCTCTGAGGCGAGCGCTGTTGAGCGGCC 1140
Db 2775 CGGAGCTGTGACATCGCAACTTCGAGCTCTCTGAGGCGAGCGCTGTTGAGCGGCC 2834
Qy 1141 CCATTACGCGGTCGCACTTCGCTGCGGGCGGCGTCAAGTGTGAGGAGCCAGAGGCGAG 1200
Db 2835 CCATTACGCGGTCGCACTTCGCTGCGGGCGGCGTCAAGTGTGAGGAGCCAGAGGCGAG 2894
Qy 1201 GTGCTCGCGCTGCAACTCAAGGCTCGGCTCGCTGTGCAAGCGCAGCGCTGAAG 1260
Db 2895 GTGCTCGCGCTGCAACTCAAGGCTCGGCTCGCTGTGCAAGCGCAGCGCTGAAG 2954
Qy 1261 GCACGGGGCCCCCTCGGCGCATGCCAGACCTTCCCGAGGGACCCCATCACAGCTCTCCAG 1320
Db 2955 GCACGGGGCCCCCTCGGCGCATGCCAGACCTTCCCGAGGGACCCCATCACAGCTCTCCAG 3014
Qy 1321 CGGACTTGGGACTGAAGGAAGAGAGGGCTCCCCCTGTGCTGGCACCCCTTAAGG 1380
Db 3015 CGGACTTGGGACTGAAGGAAGAGAGGGCTCCCCCTGTGCTGGCACCCCTTAAGG 3074
Qy 1381 GGGGCGAGCGAGCGGTGGAGCGGCACAGCCCAACCGGCACCAAGCTGCTGGGAAGGTG 1440
Db 3075 GGGGCGAGCGAGCGGTGGAGCGGCACAGCCCAACCGGCACCAAGCTGCTGGGAAGGTG 3134
Qy 1441 GAGCTCCCGGCTGCGCCCTGTGTCACGCTGGACAGTCAAGGCCCCCAACACCGGTGCCCC 1500
Db 3135 GAGCTCCCGGCTGCGCCCTGTGTCACGCTGGACAGTCAAGGCCCCCAACACCGGTGCCCC 3194
Qy 1501 GTAGCCAGCCGCCCAAGCTGGGATCCAAACCAAGCTAAATGCGCACCCGCTGGAGCCT 1560
Db 3195 GTAGCCAGCCGCCCAAGCTGGGATCCAAACCAAGCTAAATGAGCACCCGCTGGAGCCT 3254
Qy 1561 GTCTCGGGGCTGAAGACAGAGAGGCGCGCTCGCTCCGCGCCGAGGAAAGGT 1620
Db 3255 GTCTCGGGGCTGAAGACAGAGAGGCGCGCTCGCTCCGCGCCGAGGAAAGGT 3314
Qy 1621 TCA 1623
Db 3315 TCA 3317

RESULT 3

ADJ93362

ID ADJ93362 standard; DNA; 1859 BP.

XX

AC ADJ93362;

XX

DT

XX 06-MAY-2004 (first entry)

Human BGS-42 protein-related DNA clone B SeqID10.

testis-specific tubulin tyrosine-ligase-like polypeptide;

BGS-42 polypeptide; cytosolic; respiratory; anti-inflammatory; anabolic; hypertensive; osteoprotective; endocrine; anti-HIV; antiparkinsonian; antiarthritic; antisthmatic; anti-HIV; antibacterial; immunosuppressive; antiseborrheic;

dermatological; tyrosine kinase modulator; gene therapy; tubulin ligase; tubulin-carboxypeptidase; cellular proliferation; reproductive disorder; testicular disorder; testicular cancer; pulmonary cancer; lung cancer; gastrointestinal disorder; colon cancer; stomach cancer; neural disorder; brain cancer; liver cancer; proliferative condition; testis; lung;

small intestine; brain; lymph tissue; infertility; Cushing's syndrome; emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease; Parkinson's disease; immunological disorder; arthritis; asthma; AIDS; sepsis; acne; Sjogren's disease; scleroderma; human; db.

Homo sapiens.

XX WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, useful for preventing, treating or ameliorating a medical condition, e.g. aberrant cellular proliferation, reproductive disorders or testicular disorders.

XX Example 4; SEQ ID NO 10; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase-like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cytostatic, respiratory

-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, antinflammatory, anabolic, hypertensive, osteopathic, neurotropic,

anti-parkinsonian, antiarthritic, antisthmatic, anti-HIV, antibacterial, immunosuppressive, antiseborrheic or dermatological activity acting as tyrosine kinase modulators. In addition, the disclosed sequences may be

used for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or

ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders,

lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for

treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a DNA clone sequence which is related to the invention.

XX SQ Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;

Query Match 92.8%; Score 1505.8; DB 12; Length 1859;

Best Local Similarity 97.0%; Pred. No. 3.2e-311;

Matches 1574; Conservative 0; Mismatches 2; Indels 47; Gaps 2;			
QY	1	ATGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGAGCAGAGCAGA	60
Db	47	ATGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGAGCAGCAGA	106
QY	61	AGCAAGCCAGGACACAGAGGAGGAGCGCGGAGCAGCGACCTTGAGCAGCAGCAGAT	120
Db	107	AGCAAGCCAGGACACAGAGGAGGAGCGCGGAGCAGCGACCTTGAGCAGCAGCAGAT	166
QY	121	GCTGAATAATCTGAGGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC	180
Db	167	GCTGAATAATCTGAGGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC	226
QY	181	AAGGTGTCCAGGCTACCTTGGGACAGTGGAGCATGAGACATCGACATCGACAGAT	240
Db	227	AAGGTGTCCAGGCTACCTTGGGACAGTGGAGCATGAGACATCGACATCGACAGAT	286
QY	241	GCCGTGGAGACCTCACTGAGGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC	300
Db	287	GCCGTGGAGACCTCACTGAGGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC	346
QY	301	GTTTCATGGCGATGCTTTCATCTCCAATTCAAGAAATTACTTTTCGCAAGTGCAGGCTCTG	360
Db	347	GTTTCA-----TGCAGGCTCTG 363	
QY	361	CTGAATAGAAATCAGCTGTGTAACCTCTCAGACGGAATTCAGGGCTCCGGAAATCTGG	420
Db	364	CTGAATAGAAATCAGCTGTGTAACCTCTCAGACGGAATTCAGGGCTCCGGAAATCTGG	423
QY	421	ATTATAAAGCCCGGCGCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCTGTGGAG	480
Db	424	ATTATAAAGCCCGGCGCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCTGTGGAG	483
QY	481	GAGATCTCGAGTGGCAGCTGCAGACCACTCTTTTCAGGGACAAAGTGGGTGTC	540
Db	484	GAGATCTCGAGTGGCAGCTGCAGACCACTCTTTTCAGGGACAAAGTGGGTGTC	543
QY	541	CAGAAGTACATCGAGACCGCGTCTCATCTGTGACACCAAGTTTCGACATCAGCAGTGG	600
Db	544	CAGAAGTACATCGAGACCGCGTCTCATCTGTGACACCAAGTTTCGACATCAGCAGTGG	603
QY	601	TTCTCTGTCAGGACTGGAACCCCTGACCATCTGGTTTACAGGAGAGTACTTTCGGG	660
Db	604	TTCTCTGTCAGGACTGGAACCCCTGACCATCTGGTTTACAGGAGAGTACTTTCGGG	663
QY	661	TTCTCAACTCAGCGCTTCTCCCTGGACCAAGCTGGACCGCCATCCACCTGTGCAACAAC	720
Db	664	TTCTCAACTCAGCGCTTCTCCCTGGACCAAGCTGGACCGCCATCCACCTGTGCAACAAC	723
QY	721	GCCGTCCAGAAGTACTGTAAGATGATGTGGGCGGACGCCCTGTGTCGCCGACACAAC	780
Db	724	GCCGTCCAGAAGTACTGTAAGATGATGTGGGCGGACGCCCTGTGTCGCCGACACAAC	783
QY	781	ATGTGACACGACACAGGTTCCAGGAGTACCTGCAGCGCCAGGCGCTGGCGGTGG	840
Db	784	ATGTGACACGACACAGGTTCCAGGAGTACCTGCAGCGCCAGGCGCTGGCGGTGG	843
QY	841	GGCAGCGCTCATCTACCGCTCATGAAGAAGCCCATCCGCCACGACATGAAGTGGCCAG	900
Db	844	GGCAGCGCTCATCTACCGCTCATGAAGAAGCCCATCCGCCACGACATGAAGTGGCCAG	903
QY	901	GACACGTGGAGCCTCGCAAGAACAGCTTTGAGCTTACGGGGCTGACTTCGCTTGGG	960
Db	904	GACACGTGGAGCCTCGCAAGAACAGCTTTGAGCTTACGGGGCTGACTTCGCTTGGG	963
QY	961	AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCCAGCCCCACCATGCAACCGTCCACG	1020
Db	964	AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCCAGCCCCACCATGCAACCGTCCACG	1023
QY	1021	CCGGTCAAGCCGACGCTGTGTGACAGAGTGCAGGAGGACACCATCAAGGTGGCGGTGGAC	1080
Db	1024	CCGGTCAAGCCGACGCTGTGTGACAGAGTGCAGGAGGACACCATCAAGGTGGCGGTGGAC	1083

QY	1081	CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGGAGGAGCGCGGTGTGAGCCGCC	1140
Db	1084	CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGG-----AGCGGTGGTTGAGCCGCC	1139
QY	1141	CCATTTCAGCGGGTCCGACCTCTCGTGGCGGGGCTCAGTGTGAGGAGAGCCAGCAGGCGAG	1200
Db	1140	CCATTTCAGCGGGTCCGACCTCTCGTGGCGGGGCTCAGTGTGAGGAGAGCCAGCAGGCGAG	1199
QY	1201	GTGCTGCCCTCTGCAACTCAAGGGCTCGGCTCTGCTGTGGACGCGCAGCCGCTGAAG	1260
Db	1200	GTGCTGCCCTCTGCAACTCAAGGGCTCGGCTCTGCTGTGGACGCGCAGCCGCTGAAG	1259
QY	1261	GCACGGGGCCCTCGGCTATGCGCATGCGCAGCCCTGCGCAGGAGCCCATCACAGCTCTCAG	1320
Db	1260	GCACGGGGCCCTCGGCTATGCGCATGCGCAGCCCTGCGCAGGAGCCCATCACAGCTCTCAG	1319
QY	1321	CGGCACTTGGGACTGAAGAGAGAGAGGGGCTCCCCCTGGCTTGTCTGGCACCCCTTAAGG	1380
Db	1320	CGGCACTTGGGACTGAAGAGAGAGAGGGGCTCCCCCTGGCTTGTCTGGCACCCCTTAAGG	1379
QY	1381	GGGGCAGCCGAGAGCGGTGGAGCCGACAGCCCAACCGCACCAAAAGCTGTGGGAAGGTG	1440
Db	1380	GGGGCAGCCGAGAGCGGTGGAGCCGACAGCCCAACCGCACCAAAAGCTGTGGGAAGGTG	1439
QY	1441	GAGCTCCGGCCCTGCGCTGTGCGCAGCTGGACAGTCAGGCGCCCAACACCGGTGTCCCC	1500
Db	1440	GAGCTCCGGCCCTGCGCTGTGCGCAGCTGGACAGTCAGGCGCCCAACACCGGTGTCCCC	1499
QY	1501	GTAGCCAGCCGCCCAAAAGCTGGATCCAAAACAGCTAAATGGCGCACCCGCTGGAGCT	1560
Db	1500	GTAGCCAGCCGCCCAAAAGCTGGATCCAAAACAGCTAAATGGCGCACCCGCTGGAGCT	1559
QY	1561	GTGCTCGGGGCTGTAAGACAGCAGAGGGCGCTGCTGCGCGCCGCGGAGGAAAAGGT	1620
Db	1560	GTGCTCGGGGCTGTAAGACAGCAGAGGGCGCTGCTGCGCGCCGCGGAGGAAAAGGT	1619
QY	1621	TCA 1623	
Db	1620	TCA 1622	

RESULT 4

ADJ93363
ID ADJ93363 standard; DNA; 3465 BP.

XX AC ADJ93363;

XX AC

DT 06-MAY-2004 (first entry)

XX Human BGS-42 protein-related DNA clone C SeqID11.

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nontropic; antiparkinsonian; antichratic; antiaethmatic;
anti-Hiv; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; ds.

XX Homo sapiens.

XX WO2004005487-A2.

XX PN

XX XX

XX PD 15-JAN-2004.

XX XX

PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
PI WPI; 2004-099381/10.
XX
DR
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Example 4; SEQ ID NO 11; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, anticiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match 90.9%; Score 1475.8; DB 12; Length 3465;
Best Local Similarity 95.7%; Pred. No. 8.9e-305;
Matches 1554; Conservative 0; Mismatches 2; Indels 67; Gaps 1;

QY 1 ATGGCATCCAGCATCCTCAAGTGGGTGGTCCAGCCACGAGAGTGCAGCAGGAGCAGCAGA 60
DB 1678 ATGGCATCCAGCATCCTCAAGTGGGTGGTCCAGCCACGAGCTGCAGCAGGAGCAGCAGA 1737

QY 61 AGCAAGCCGAGGACACAGAGGAGGAGGCGCGGAGCAGCGACTGCAGCAGGAGGCGCAAGAT 120
DB 1738 AGCAAGCCGAGGACACAGAGGAGGAGGCGCGGAGCAGCGACTGCAGCAGGAGGCGCAAGAT 1797

QY 121 GCTGAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGCGCTTGTGGACATCGCGTGC 180
DB 1798 GCTGAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGCGCGCTTGTGGACATCGCGTGC 1857

QY 181 AAGGTGTGTCAGGCGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 240
DB 1858 AAGGTGTGTCAGGCGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 1917

QY 241 GCCGTGGAGGACCTCACTAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
DB 1918 GCCGTGGAGGACCTCACTAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977

QY 301 GTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGGCCAGGCTCTG 360
DB 1978 GTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGGCCAGGCTCTG 2037

QY 361 CTGAATAGAAATCACGCTCTGTGAAACCTCTCAGACGGACATTTGACGGGCTCCGGAAACATCTGTG 420
DB 2038 CTGAATAGAAATCACGCTCTGTGAAACCTCTCAGACGGACATTTGACGGGCTCCGGAAACATCTGTG 2097

QY 421 ATTATTAAGCCCGCGGCGCAAGTCCCGGGGCGGAGACATAGTGTGCATGAGACCGTGTGGAG 480
DB 2098 ATTATTAAGCCCGCGGCGCAAGTCCCGGGGCGG----- 2129

QY 481 GAGATCCTGGAGCTGCGACCTGCAGACCACCTCTTTTCAGGAGCAACAAGTGGGTGGTC 540
DB 2130 -----AGGACAAACAAGTGGGTGGTC 2150

QY 541 CAGAAGTACATCEGAGACGCCGTGTCTCATCTGTGACACCAAGTTCCACATCAGACAGTGG 600
DB 2151 CAGAAGTACATCEGAGACGCCGTGTCTCATCTGTGACACCAAGTTCCACATCAGACAGTGG 2210

QY 601 TTCTCCTGTCACGACCTGGAAACCCCTTGACCATCTGTTCTTACAGGAGATTTACTTTGCGG 660
DB 2211 TTCTCCTGTCACGACCTGGAAACCCCTTGACCATCTGTTCTTACAGGAGATTTACTTTGCGG 2270

QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCCACCTGTGCACAAAC 720
DB 2271 TTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCCACCTGTGCACAAAC 2330

QY 721 GCCGTCCAGAAGTACTGTAAGAAATGATGTGGCCGAGCCCCCTGTCTGCCCGCACAAAC 780
DB 2331 GCCGTCCAGAAGTACTGTAAGAAATGATGTGGCCGAGCCCCCTGTCTGCCCGCACAAAC 2390

QY 781 ATGTGGACACGACACCAAGTTCCAGGAGTACTTGACGCGCAGCGGCGCTGCCCGCTGTGG 840
DB 2391 ATGTGGACACGACACCAAGTTCCAGGAGTACTTGACGCGCAGCGGCGCTGCCCGCTGTGG 2450

QY 841 GCGAGCGTCTATCTACCCGTCATGAAGAGGCCATCGCCACGCGCCATGAAGTGGCGCCAG 900
DB 2451 GCGAGCGTCTATCTACCCGTCATGAAGAGGCCATCGCCACGCGCCATGAAGTGGCGCCAG 2510

QY 901 GACACGCTGAGCCTCGCAAGAACAGCTTTGAGCTTACGCGGCTGACTTCGTCCTTTGGG 960
DB 2511 GACACGCTGAGCCTCGCAAGAACAGCTTTGAGCTTACGCGGCTGACTTCGTCCTTTGGG 2570

QY 961 AGGACCTTCAGGCGCTTGGCTGATCGAGATCAATTCAGCGCCCGCCACCATGACCGCTCCAG 1020
DB 2571 AGGACCTTCAGGCGCTTGGCTGATCGAGATCAATTCAGCGCCCGCCACCATGACCGCTCCAG 2630

QY 1021 CCGGTCACGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGTGGCGCTGGAC 1080
DB 2631 CCGGTCACGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGTGGCGCTGGAC 2690

QY 1081 CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGTGAGGAGCGCGGTGTTGAGCGCGCC 1140
DB 2691 CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGTGAGGAGCGCGGTGTTGAGCGCGCC 2750

QY 1141 CCATTCAGCGGCTCCGACCTCTGCGTGGCGGCGCTCAGTGTGAGGAGAGCCAGGAGGCGAG 1200
DB 2751 CCATTCAGCGGCTCCGACCTCTGCGTGGCGGCGCTCAGTGTGAGGAGAGCCAGGAGGCGAG 2810

QY 1201 GTGCTGCCGCTGTCAAACCTCAAGGCTCGGCTCGCTGTGGACGCGCAGCGCTGAAAG 1260
DB 2811 GTGCTGCCGCTGTCAAACCTCAAGGCTCGGCTCGCTGTGGACGCGCAGCGCTGAAAG 2870

QY 1261 GCACGCGGCGCTTCGCGCCATGCCAGACCTTGCCAGAGGACCCCGCATCAACAGCTCTCCAG 1320
DB 2871 GCACGCGGCGCTTCGCGCCATGCCAGACCTTGCCAGAGGACCCCGCATCAACAGCTCTCCAG 2930

QY 1321 CGGGACTGGGACTGAAGGAAGAGGGGCTCCCGCTGGCGCTTGTGCGACCCCTTAAGG 1380
DB 2931 CGGGACTGGGACTGAAGGAAGAGGGGCTCCCGCTGGCGCTTGTGCGACCCCTTAAGG 2990

QY 1381 GGGGACCGCGAGAGCGGTGGAGCGCGCACAGCCCAACCCGACCAAAAGCTGTCTGGGAAGGTG 1440
DB 2991 GGGGACCGCGAGAGCGGTGGAGCGCGCACAGCCCAACCCGACCAAAAGCTGTCTGGGAAGGTG 3050

QY 1441 GAGCTCCGGCTCCCTGTCGCCACCTGACAGTCAGGCCCAAAACACCGGTGTCCTCC 1500
| | | | |
Db 3051 GAGCTCCGGCTCCCTGTCGCCACCTGACAGTCAGGCCCAAAACACCGGTGTCCTCC 3110
| | | | |
QY 1501 GTAGCCAGCCCGCCAAAGCTGGATCCAAACAGCTAAATGGCGCACCCGCTGGAGCCT 1560
| | | | |
Db 3111 GTAGCCAGCCCGCCAAAGCTGGATCCAAACAGCTAAATGGCGCACCCGCTGGAGCCT 3170
| | | | |
QY 1561 GTGCTGGGGCTGAGACAGCAGAGGGCGCTGGCTGGCGCCCGCGAGGAAAGGT 1620
| | | | |
Db 3171 GTGCTGGAGCCTGAGACAGCAGAGGGCGCTGGCTGGCGCCCGCGAGGAAAGGT 3230
| | | | |
QY 1621 TCA 1623
| | |
Db 3231 TCA 3233
| | |
RESULT 5
ADJ93361
ID ADJ93361 standard; DNA; 1939 BP.
AC ADJ93361;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone A SeqID9.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory; respiratory; gastrointestinal; neuroprotective; endocrine; anti-inflammatory; anabolic; hypertensive; osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic; anti-HIV; antibacterial; immunosuppressive; antiseborrheic; dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase; tubulin-carboxypeptidase; cellular proliferation; reproductive disorder; testicular disorder; testicular cancer; pulmonary disorder; lung cancer; gastrointestinal disorder; colon cancer; stomach cancer; neural disorder; brain cancer; liver cancer; proliferative condition; testis; lung; small intestine; brain; lymph tissue; infertility; Cushing's syndrome; emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease; Parkinson's disease; immunological disorder; arthritis; asthma; AIDS; sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX Homo sapiens.
XX WO2004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX PA
XX Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, useful for preventing, treating or ameliorating a medical condition, e.g. aberrant cellular proliferation, reproductive disorders or testicular disorders.
XX
XX
XX Example 4; SEQ ID NO 9; 343bp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase-like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cytosolic, respiratory -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, anti-inflammatory, anabolic, hypertensive, osteopathic, nontropic, antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial, immunosuppressive, antiseborrheic or dermatological activity acting as tyrosine ligase modulators. In addition, the disclosed sequences may be

CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a DNA clone sequence which is related to the invention.

XX
SQ Sequence 1939 BP; 421 A; 592 C; 589 G; 337 T; 0 U; 0 Other;

Query Match 90.8%; Score 1474.2; DB 12; Length 1939;
Best Local Similarity 95.7%; Pred. No. 1.7e-304;
Matches 1553; Conservative 0; Mismatches 3; Indels 67; Gaps 1;

QY 1 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGAGCAGAGA 60
| | | | |
Db 153 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGAGCAGAGA 212
| | | | |
QY 61 AGCAAGCCAGGACACAGAGGGAGGAGCGGAGCAGCAGCCTTGAGCAGCAGCAAGAT 120
| | | | |
Db 213 AGCAAGCCAGGACACAGAGGGAGGAGCGGAGCAGCAGCCTTGAGCAGCAGCAAGAT 272
| | | | |
QY 121 GCTGAAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCCGCTGC 180
| | | | |
Db 273 GCTGAAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCCGCTGC 332
| | | | |
QY 181 AAGTGTGCCAGCCTTACCTGGGGAGCTGGAGCATGAGCAGCATCGACAGTTCAGCAGAT 240
| | | | |
Db 333 AAGTGTGCCAGCCTTACCTGGGGAGCTGGAGCATGAGCAGCATCGACAGTTCAGCAGAT 392
| | | | |
QY 241 GCGTGGAGGACCTCACTGAGGCGGAGTGGGAGGAGCTGAGCAGCAGTACTACTCCCTC 300
| | | | |
Db 393 GCGTGGAGGACCTCACTGAGGCGGAGTGGGAGGAGCTGAGCAGCAGTACTACTCCCTC 452
| | | | |
QY 301 GTTCATGGCGATGCTTTTCATCTCAATTCAGAAATTTACTTTTCGAGTGCAGGCTCTG 360
| | | | |
Db 453 GTTCATGGCGATGCTTTTCATCTCAATTCAGAAATTTACTTTTCGAGTGCAGGCTCTG 512
| | | | |
QY 361 CTGAATAGATACAGTCTGTGAACCTTCAGACGAGCATTTGACGGGCTCCGGAACATCTGG 420
| | | | |
Db 513 CTGAATAGATACAGTCTGTGAACCTTCAGACGAGCATTTGACGGGCTCCGGAACATCTGG 572
| | | | |
QY 421 ATTATAAGCCCGCGCCCAAGTCCCGGGGCGAGACATAGTGTGATGGACCGTGTGGAG 480
| | | | |
Db 573 ATTATAAGCCCGCGCCCAAGTCCCGGGGCGG----- 604
| | | | |
QY 481 GAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTTCAGGGGACAAAGTGGGTGCTC 540
| | | | |
Db 605 -----AGGACAAACAAAGTGGGTGCTC 625
| | | | |
QY 541 CAGAAGTACATCAGACGGCGTGTGTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG 600
| | | | |
Db 626 CAGAAGTACATCAGACGGCGTGTGTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG 685
| | | | |
QY 601 TTCTCTCGTCCGAGCTGGAAACCCCTTGACCATCTGTTCTACAGGAGAGTACTTTCGG 660
| | | | |
Db 686 TTCTCTCGTCCGAGCTGGAAACCCCTTGACCATCTGTTCTACAGGAGAGTACTTTCGG 745
| | | | |
QY 661 TTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGACAGCGCCCATCCACCTGTGCAACAC 720
| | | | |
Db 746 TTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGACAGCGCCCATCCACCTGTGCAACAC 805
| | | | |

XX
SQ Sequence 726 BP; 157 A; 227 C; 214 G; 128 T; 0 U; 0 Other;
Query Match 44.7%; Score 726; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 4.7e-145;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 397 ATTGACGGGCTCCGGAACATCTGGATTATATAAGCCCGGCCCAAGTCCCGGGGCGCGAGC 456
DB 1 ATTGACGGGCTCCGGAACATCTGGATTATATAAGCCCGGCCCAAGTCCCGGGGCGCGAGC 60
QY 457 ATAGTGTGCATGACACCGCTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTT 516
DB 61 ATAGTGTGCATGACACCGCTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTT 120
QY 517 TCAGGAGCAACAAAGTGGGTGTCAGAGTACATCAGAGACGCGCTGCTCATCTGTGAC 576
DB 121 TCAGGAGCAACAAAGTGGGTGTCAGAGTACATCAGAGACGCGCTGCTCATCTGTGAC 180
QY 577 ACCAAGTTCACATCAGACAGTGGTTCCTCGTCAAGGACTGGAAACCCCTGACCATCTGG 636
DB 181 ACCAAGTTCACATCAGACAGTGGTTCCTCGTCAAGGACTGGAAACCCCTGACCATCTGG 240
QY 637 TTCTACAAGGAGAGTACTTTCGCGGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 696
DB 241 TTCTACAAGGAGAGTACTTTCGCGGTTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 300
QY 697 AGCGCCATCACCCTGTGCAACAAACGCGCTCCAGAGTACCTGAAGATGATGTGGGCGC 756
DB 301 AGCGCCATCACCCTGTGCAACAAACGCGCTCCAGAGTACCTGAAGATGATGTGGGCGC 360
QY 757 AGCCCTCTGCTGCCGACACACATGTCAGACGACGACGACGACGACGACGACGACGAC 816
DB 361 AGCCCTCTGCTGCCGACACACATGTCAGACGACGACGACGACGACGACGACGACGAC 420
QY 817 CGCCAGGCGCGTGGCGCGTGTGGGCGAGCTCATCTACCCGCTCCATGAAGAAGGCCATC 876
DB 421 CGCCAGGCGCGTGGCGCGTGTGGGCGAGCTCATCTACCCGCTCCATGAAGAAGGCCATC 480
QY 877 GCCCAGCCATGAAGTGGGCCAGGACCACTGGAGCTCCGCAAGAACAGCTTTGAGCTC 936
DB 481 GCCCAGCCATGAAGTGGGCCAGGACCACTGGAGCTCCGCAAGAACAGCTTTGAGCTC 540
QY 937 TAGCGGGCTGACTTCGCTCTTGGAGGGGACTTCAGGCGCTGGCTGATCGAGATCAATTC 996
DB 541 TAGCGGGCTGACTTCGCTCTTGGAGGGGACTTCAGGCGCTGGCTGATCGAGATCAATTC 600
QY 997 AGCCCAACCATGACACCGCTCCAGCGCGTCCAGCGCCAGCTGTGTGCACAGGTGCAGGAG 1056
DB 601 AGCCCAACCATGACACCGCTCCAGCGCGTCCAGCGCCAGCTGTGTGCACAGGTGCAGGAG 660
QY 1057 GACACCATCAAGTGGCGGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 1116
DB 661 GACACCATCAAGTGGCGGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 720
QY 1117 AGGCAG 1122
DB 721 AGGCAG 726

RESULT 7
ADQ17814/C
ID ADQ17814 standard; DNA; 101270 BP.
XX
AC ADQ17814;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 631.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX
ds.

OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PP 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 631; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytotatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;
Query Match 30.9%; Score 501; DB 12; Length 101270;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1120 CAGCCGGTGTGAGCGCGCCCAATTCAGCGGTCCGACCTCTCGTGGCGGGCGTCAGT 1179
DB 43637 CAGCCGGTGTGAGCGCGCCCAATTCAGCGGTCCGACCTCTCGTGGCGGGCGTCAGT 43578
QY 1180 GTGAGGAGAGCCAGGAGGAGGTGTGCGCTGCAACCTCAAGCCTCGGCTCGCTG 1239
DB 43577 GTGAGGAGAGCCAGGAGGAGGTGTGCGCTGCAACCTCAAGCCTCGGCTCGCTG 43518
QY 1240 TTGGACGCGCAGCGCTGAAGGCACGGGGCCCTCGGCCATGCGACACCTGCCAGGGA 1299
DB 43517 TTGGACGCGCAGCGCTGAAGGCACGGGGCCCTCGGCCATGCGACACCTGCCAGGGA 43458
QY 1300 CCCCACATCAACAGCTCTCCAGCGGGACTTGGAGCTGAAGAGAGAGAGAGAGAGAGAG 1359
DB 43457 CCCCACATCAACAGCTCTCCAGCGGGACTTGGAGCTGAAGAGAGAGAGAGAGAGAG 43398
QY 1360 GCCTTGTGTCACCCCTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCGC 1419
DB 43397 GCCTTGTGTCACCCCTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCGC 43338
QY 1420 ACCAAAGCTGTGGGAAGGTGGAGCTCCCGGGCTGCCCTGTGCCACGCTGGACAGTCAG 1479
DB 43337 ACCAAAGCTGTGGGAAGGTGGAGCTCCCGGGCTGCCCTGTGCCACGCTGGACAGTCAG 43278
QY 1480 GCCCCAAAACACCGGTGTCCCGTAGCCCGCCGCAAGCTGGGATCAAAACAGCTA 1539
DB 43277 GCCCCAAAACACCGGTGTCCCGTAGCCCGCCGCAAGCTGGGATCAAAACAGCTA 43218
QY 1540 AATGCGACCCGCTGGAGGCTGTGCTGCGGGGCTGAAGCAGCAGAGGCGCGCTGCGT 1599

Db 43217 AATGGCACCCGCTGGAGCCTGTGCTGCGGGCCTGAAGACAGCAGAGGCGCGCTGCGT 43158

QY 1600 CCGCGCGCCGGAGGAAAGGT 1620

Db |||||

Db 43157 CCGCGCGCCGGAGGAAAGGT 43137

RESULT 8

AAS99894

XX AAS99894 standard; cDNA; 2380 BP.

AC AAS99894;

XX

XX 12-MAR-2002 (first entry)

XX

DE Polynucleotide encoding human cytoskeleton-associated protein (CYSKP) #5.

XX

KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;

KW cell proliferative disorder; inflammatory disorder; prion disease;

KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;

KW neurological disorder; cell motility disorder; reproductive disorder;

KW spinal cord disease; central nervous system disorder; mental disorder;

KW gene therapy; cancer.

XX

XX Homo sapiens.

OS

XX

XX WO200185942-A2.

XX

XX 15-NOV-2001.

XX

XX 03-MAY-2001; 2001WO-US014355.

XX

XX 05-MAY-2000; 2000US-0201960P.

PR 08-MAY-2000; 2000US-0202729P.

PR 05-JUN-2000; 2000US-0209705P.

PR 07-JUN-2000; 2000US-0210149P.

PR 21-JUN-2000; 2000US-0213215P.

XX

XX (INCY-) INCYTE GENOMICS INC.

XX

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;

PI Azinzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;

PI Policky JL;

XX

DR WPI; 2002-062248/08.

DR P-PSDB; AAU74334.

XX

PT New cytoskeleton-associated proteins and polynucleotides, useful for

PT diagnosing, preventing and treating cell proliferative, autoimmune,

PT inflammatory, neurological, cell motility, reproductive and muscle

PT disorders.

XX

PS Claim 5; Page 171; 194pp; English.

XX

XX The invention relates to human cytoskeleton-associated polypeptides

CC (CYSKP) and their associated polynucleotide sequences. The sequences are

CC useful in the treatment of disorders associated with overexpression or

CC underexpression of CYSKP in a patient. The disorders include cell

CC proliferative disorders (such as cancer, actinic keratosis,

CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),

CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,

CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus

CC and anaemia), vesicle trafficking disorders (such as

CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),

CC gastrointestinal disorders, prion disease, neurological disorders (such

CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,

CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis

CC and other motor neuron disorders), cell motility disorders, reproductive

CC disorders (such as endometriosis and polycystic ovary syndrome), muscle

CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,

CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord

CC diseases, central nervous system disorders (such as Down syndrome and

CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

CC Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP

CC of the invention

XX

SQ Sequence 2380 BP; 580 A; 685 C; 670 G; 445 T; 0 U; 0 Other;

Query Match 26.7%; Score 432.8; DB 6; Length 2380;

Best Local Similarity 59.1%; Pred. No. 2e-82;

Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;

QY 108 CAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGTCTAG-GGGCCCTCCGGGGCAGCTTG 166

Db |||||

Db 361 CAGGAGACAGCAGCCCAAGAAACAGGAGAAAACCCAGTGTGGTGTCCCCAGAGTTG 420

QY 167 TGGACATCGCGTGCACAAAGGTGTCAGAGGCTACTCTGGGGCAGCTGAGGACATGAGACATCG 226

Db |||||

Db 421 TGGATGAAGCTCTGTGTGCTGCGAGGAGTACTTAGCAACTTGGCCCAATGGACATCG 480

QY 227 ACACCTCAGCAGATCCGCTGGAGGACCTCACTAGGCCGAGTGGGAGGACCTGACCCAGC 286

Db |||||

Db 481 ACAAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGGCTGGTCCCTCTTCCTCCAGC 540

QY 287 AGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAAGAAATTAATTTTCG 346

Db |||||

Db 541 GCTACTACCAAGTGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCAGTCCAGC 600

QY 347 AGTGCCAGGCTCTGCTGAATAGAAATCACTGTGTGAACCTCTCAGACGGAATGACGGGC 406

Db |||||

Db 601 GCTGTGAGGACATCTCTGCAGCAGCTGCAGGCCGTGGTACCCAGATAGACATGGAGGG 660

QY 407 TCCGGAACATCTGGATTATAAGCCGCGGCCCAAGTCCCGGGCCGAGACATAGTGTGCA 466

Db |||||

Db 661 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGTGAGGAGCATCATGTGCA 720

QY 467 TGGACCGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGACA 526

Db |||||

Db 721 TGGACCACTGGAGGAGATGCTGAAGCTGTGAACGGCAACCCCGTGGTGATGAAGGACG 780

QY 527 ACAAGTGGGTGTCAGAAAGTACATCGAGACGCGCTGCTCATCTGTGTGACACCAAGTTG 586

Db |||||

Db 781 GCAAGTGGGTGTCAGAAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAAGTTG 840

QY 587 ACATCAGACAGTGGTTCCTCGTCCGACTGGAACCCCTGACCATCTCGTTCTACAAGS 646

Db |||||

Db 841 ACCTCAGACAGTGGTTCCTGGTAACTGACTGGAACCCACTTACCGTGTGGTCTACCGGG 900

QY 647 AGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCC 706

Db |||||

Db 901 ACAGCTATATCGCTTTTCCACGCGAGCCCTTCTCCCTGAAGAACCTGGACAACTCAGTGC 960

QY 707 ACTGTGCAACAAACGCGCTCCAGAGTACCTGAAGAAATGATGTGGGCCGCGAGCCCTGCG 766

Db |||||

Db 961 ACCTGTGCAACAACTCCATCCAGAAAGCACCTGGAGAACTCATGCCCATCGGCATCCACTGC 1020

QY 767 TGC CGCGCACACAACTGTGGACCAACAGCTTCCAGGAGTACCTGACGCGCCAGCGGC 826

Db |||||

Db 1021 TTCCGCGCAGACAACTATGTTGTTAGCAGAGGTTCAGGCCCACTTCGAGGAGATGGGTG 1080

QY 827 GTGGCGCGCTGTGGGGCAGCGTCACTTACCCGCTCCATGAAGAGGCCATCGCCACGCCA 886

Db |||||

Db 1081 CCCCAATGCTTGGTCCACCATCATCGTCCCTGGCATGAAGGATGCTGTGATCCACGCAC 1140

QY 887 TGAAGGTGGCCCGCAGGACCACTGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTG 946

Db |||||

Db 1141 TTCAAGACCTCCCGAGGACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTATGGCGCTG 1200

QY 947 ACTTGTCTCTGGGAGGAGCTTTCAGCCCTGCTGATCGAGATCAATTTCCAGCCCCACCA 1006

Db |||||

Db 1201 ACTTGTGTTCGGGGAGGAGCTTCCAGCCCTGCTGATTCAGATCAACGCCAGCCCCACCA 1260

QY 1007 TGCACCCGCTCCAGCGCGGTTCACGGGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCA 1066

Db |||||

Db 1261 TGGCACCTCCACAGCAGTCACTGCGCCGCTCTGTGCTGGCGTGCAGAGCTGACACCTGCG 1320

QY 1067 AGTGGCC-----GTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGT 1114
Db |||||
QY 1321 CGTGTGTCATTGACCGGATGCTGACCGCAACTGTGACAGGAGCTTTGAGCTCACT 1380
Db |||||
QY 1115 GAGGAGCGCGTGTGAGCGCGCCCACTTCAGCGGGTCCGACCTCTGGTGTGCGGGCG 1174
Db |||||
QY 1381 ATAAGCAGCTGTGTGAGGTGCTCAATATGTGGGCATCCGGCTCTCTGTAGAGGCT 1440
Db |||||
QY 1175 TCAGTGTGAGAGCAGGAGGAGGTGCTGCCGTCTGCAACCTCAAGGCTCTGGCCT 1234
Db |||||
QY 1441 TCACCATCAAGAAGCCATCGCATGTGTCTATCGGCGGATGGGGTCCGCCAGCATCC 1500
Db |||||
QY 1235 CGCTGTGGACGCGCAGCGCTGAAGCAGCGGGCCCTCGGCCATGCC-AGACCCCTGCC 1293
Db |||||
QY 1501 CTCTGTGACCCAGGAGGCTCTGGGGAAGCAAGGACTCGGGATCCCTACCCACAGGT 1560
Db |||||
QY 1294 CAGGAGCCCCATCACCAGCTCTCCAGCGGACTTGGGACTGAAGGAAGAAGGGGCTC 1353
Db |||||
QY 1561 CAGCTTCTAGGAAAGCACTGGGSCCAGGAGCTTGGGGCAGGTGAGAGCCAGTCTCCA 1620
Db |||||
QY 1354 CCCCTGCTTGTGCGCACCTTTAAGGGGGGAGCCGAGAGCGGTGAGCGGCACAGCCC 1413
Db |||||
QY 1621 CTGCGCACCTCTCAGCCCCCGGAAGGGGAAGCAAGGCGCAAAAGGGCCACAGCCC 1680
Db |||||
QY 1414 ACCCGCACAAAGCTGTCTGGGAAGTGTGAGCTCCCGC 1451
Db |||||
QY 1681 TGGTCTGCCCAATCTCTGGGAGTGGATGCCCCCAGC 1718
Db |||||
RESULT 9
AAH16735
ID AAH16735 standard; cDNA; 2326 BP.
XX
AC AAH16735;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15920.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
XX
PR 27-AUG-1999; 99JP-00300253.
XX
PR 11-JAN-2000; 2000JP-00118776.
XX
PR 02-MAY-2000; 2000JP-00183767.
XX
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 15920; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH58893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2326 BP; 524 A; 701 C; 642 G; 459 T; 0 U; 0 Other;

Query Match 26.6%; Score 431.6; DB 4; Length 2326;
Best Local Similarity 63.5%; Pred. No. 3.5e-82;
Matches 698; Conservative 0; Mismatches 389; Indels 13; Gaps 2;

QY 108 CAGCAGCGAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCTCCCGGGCAGCTTG 166
Db |||||
QY 620 CAGGAGACAAGCAGCGCCCAAGAAAACAGAGAAAAACCCAGTGTGGTGTCCCCCAGAGTTG 679
Db |||||
QY 167 TGGACATCGCGTCAAGGTGTGCCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCG 226
Db |||||
QY 680 TGGATGAAGCTCTGTGTGCTGCGAGAGTACTTTAGCAACTTGGCCCAATGAGACATCG 739
Db |||||
QY 227 ACACGTGACGAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACCCAGC 286
Db |||||
QY 740 ACAAGGACCTGGAGGCCCCGCTGCTACCTACCCCGAGGGCTGCTCTCTCTCCAGC 799
Db |||||
QY 287 AGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAATTACTTTTCG 346
Db |||||
QY 800 GCTACTACTCAAGTGGTCCAGAAAGGGGCGAGAACTCAGGCACCTCGACACTCAGGTCCAGC 859
Db |||||
QY 347 AGTGCCAGGCTCTGTGAAATAGAAATCACGTCTGTGAACCTCAGAGCCATGAGACGATGACGGGC 406
Db |||||
QY 860 GCTGTGAGGACATCTCTGACGAGCTGAGGCCGTGTGTACCCAGATAGACATGGAAGGGG 919
Db |||||
QY 407 TCCGGAACATCTGGATTATAAAGCCCGGCCCAAGTCCCGGGCCGAGACATAGTGTGCA 466
Db |||||
QY 920 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGCGAGGACATCATGTGCA 979
Db |||||
QY 467 TGGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGAGACCAACCTCTTTCCAGGGACA 526
Db |||||
QY 980 TGGACCACTTGGAGGAGATGCTGAAGCTGGTGAACCGCAACCCCGTGTGTATGAAGGAGC 1039
Db |||||
QY 527 ACAAGTGGTGTGTCAGAAAGTACATCGAGACGCGGTGCTCATCTGTGACACCAAGTTTCG 586
Db |||||
QY 1040 GCAAGTGGTGTGTCAGAAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAGTTTG 1099
Db |||||
QY 587 ACATCAGACAGTGGTTCTCTCGTCACGAGCTGGAAACCCCTCGACCATCTGTTCTTACAAGG 646
Db |||||
QY 1100 ACCTCAGACAGTGGTTCTCTGTTAACTGACTGGAACCCACTTACCGTGTGTCTTACCGCG 1159
Db |||||
QY 647 AGATTACTTTCGGTTCTCAACTCAGGGTTCTCTCTGGACAAGCTGGACAGCGCCATCC 706
Db |||||
QY 1160 ACAGCTATATCCCTTTTCCACGACGCTTCTCTCTGGAAGAACCTTGGAACAACCTCAGTGC 1219
Db |||||
QY 707 ACCTGTGCAACACGCGCTCCAGAACTCTGAAGATGATGTGGGCCGAGCCCTCTGC 766
Db |||||
QY 1220 ACCTGTGCAACAACTCCATCCAGAACCTCTGAGAACTCATGCGATCGGCATCCACTGC 1279
Db |||||
QY 767 TGCCCCCACAACAATGTGGACCCAGACCCAGGTTCCAGGAGTACCTGCAGCGCAGGGCC 826
Db |||||
QY 1280 TTCCGCCAGACAACATGTGTCTAGCCAGAGGTTCCAGGGCCCACTCTGCAGGAGTGGGTG 1339
Db |||||

AC	AD573168;	QY	140	AGCTCAG-GGGCCCTCCCGGGGCGAGTTGTGGACATCGCGTGAAGGTGTGCGAGSCCTAC	198
XX					
DT	18-NOV-2004 (first entry)	Db	814	AACCCAGTGTGGTGTCCCGAGAGTTGTGGATGAAGCTGTGTGTGGTGGAGAGTAC	873
XX					
DE	Human kidney tumour specific cDNA, SEQ ID 1765.	QY	199	CTGGGGCAGCTGGAGCATGAGGACATCGACACAGTCAGCAGATGCGGTGGAGGACCTCACT	258
XX					
KW	Human; ss; kidney tumour; kidney cancer; cytostatic; gene therapy;	Db	874	CTTAGCAACTTGGCCACATGGACATCGACAGGACCTGGAGGCCCGCTGTACTCTCACC	933
KW	T-cell; immune response.				
XX		QY	259	GAGCCGAGTGGGAGGACCTGACCCAGAGATCTACTCCCTCGTTTCATGCGGATGCTTTTC	318
XX					
OS	Homo sapiens.	Db	934	CCGAGGGCTGGTCCCTCTCTCCAGCGCTACTACCAAGTGGTCCAGGAAGGGGCGAGAA	993
XX					
XX	US2003109434-A1.	QY	319	ATCTCCAAATTCAGAAATTAATTTTCGAGTGGCCAGGCTCTGCTGAATAGATCACTGCT	378
XX					
XX	12-JUN-2003.	Db	994	CTCAGGCACCTCGACACTCAGGCTCCAGCGCTGTGGAGACATCTCTGCGAGCAGCTGAGGCC	1053
XX					
XX	19-MAR-2002; 2002US-00102524.	QY	379	GTGAACCCCTCAGACGGACATTTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCC	438
XX					
PR	19-MAR-2001; 2001US-0277245P.	Db	1054	GTGGTACCCAGATAGACATGGAAGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCC	1113
PR	21-DEC-2001; 2001US-0343340P.				
XX		QY	439	AAGTCCCGGGCCGAGACATAGTGTGATCGAACCGTGTGAGGAGATCTCTGGAGCTGCA	498
XX					
PA	(CORI-) CORIXA CORP.	Db	1114	AAGTCCCGCGAGGAGGCATCATGTGCATGAGACCACTCGAGGAGATGCTGAAGTGGTG	1173
XX					
PI	Algate PA, Mannion J, Gaiger A, Gordon B, Harlocker SL;	QY	499	GCTGCGAGACCACTCTTTCCAGGGCAACAAGTGGGTGCTCCAGAGATACATCGAGACG	558
XX					
DR	WPI; 2002-759855/82.	Db	1174	AACGGCAACCCCGTGGTGATGAAGCGGCAAGTGGGTGGTGCAGAGATATATTTAGCGG	1233
XX					
PT	New isolated polynucleotides and polypeptides, useful for detecting the	QY	559	CCGCTGCTCATCTGTGACACAAATTCGACATCAGACAGTGGTTCCTCGTCAGCAGACTGG	618
PT	presence of, and treating cancer, particularly kidney cancer by				
PT	stimulating T-cells specific for a tumor protein, and stimulating immune	Db	1234	CCCTCTCATCTTTGGCACCAAGTTTGACTCAGACAGTGGTTCCTGTAACCTGACTGG	1293
XX	response in a patient.				
PS	Claim 1; SEQ ID NO 1765; 78pp; English.	QY	619	AACCCCTGACCATCTGGTTTCTCAAGGAGATTAATTGGCGTTCTCAACTCAGCGCTTC	678
XX					
CC	The invention relates to a new isolated polynucleotide (a Human kidney	Db	1294	AACCCCTGACCATCTGGTTTCTCAAGGAGATTAATTGGCGTTCTCAACTCAGCGCTTC	1353
CC	tumour specific cDNA) comprising any one of the 1855 sequences identified				
CC	in the specification (or their complements, degenerate variants,	QY	679	TCCTTGCAAGCTGGACAGCGGCATCCACTGTGCAACACCGCGTGCAGAGTACTCG	738
CC	sequences consisting of at least 20 contiguous residues them, sequences				
CC	that hybridise to them under highly stringent conditions or sequences	Db	1354	TCCTTGCAAGCTGGACAACTCAGTGACCTGTGCAACAACTCCATCCAGAGACACTG	1413
CC	having at least 75 or 90% sequence identity to the 1855 sequences. Also				
CC	included are detecting/determining the presence of cancer in a patient,	QY	739	AAGAATGATGTGGCGCGCAGCCCCCTGCTGCCGCGCACAAACATGTGGACAGCAGCAGG	798
CC	stimulating an immune response in a patient; treating kidney cancer in a				
CC	patient, an isolated polypeptide encoded by one of the 1855 sequences, an	Db	1414	GAGAACTCATGCCATCGGCATCCACTGCTTCGCCAGACAACTGTGGTCTAGCCAGAGG	1473
CC	expression vector comprising the polynucleotide operably linked to an				
CC	expression control sequence, a host cell transformed/transfected with the	QY	799	TTCCAGAGTACTCTGACGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTCATCTACCCG	858
CC	vector, an isolated antibody (or its antigen-binding fragment) that				
CC	specifically binds to the protein, a fusion protein comprising at least	Db	1474	TTCCAGGCGCCACCTGCGAGGAGATGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1533
CC	one the protein, stimulating and/or expanding T-cells specific for a	QY	859	TCCATGAAGAGGCCATCGCCCGCCATGAAGTGGCGCCAGGACAGCAGCAGCAGCAGCAGG	918
CC	tumour protein, an isolated T-cell population comprising the T-cells, a				
CC	composition comprising a first component (such as a carrier or	Db	1534	GGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGG	1593
CC	immunostimulant) and a second component (comprising one of the				
CC	polynucleotides, the polypeptides, an antibody, T-cell or an antigen-	QY	919	AAGAACAGCTTTGAGCTCTACGGGGTGTACTTCTCTCTGGGAGGAGACTTTCAGGCGCTGG	978
CC	presenting cell that expresses the polynucleotide) and a diagnostic kit				
CC	comprising at least one of the oligonucleotides, or at least one antibody	Db	1594	AAGGCCAGCTTTGAGCTCTATGGCGTGTACTTGTGTGGGGGAGACTTCCAGGCGCTGG	1653
CC	and a detection reagent comprising a reporter group. The polynucleotides,				
CC	polypeptides, antibodies and antigen-presenting cells are useful for	QY	979	CTGATCGAGATCAATTCAGCCCCCAATGCAACCGTTCACCGCGGTTCAGGCGCCAGCTG	1038
CC	detecting the presence of, and treating cancer, particularly kidney				
CC	cancer by stimulating and/or expanding T-cells specific for a tumour	Db	1654	CTGATTTGAGATCAACGCGAGCCCCCGATGGCACCTCCACAGCAGTCACTCTCCCGGCTC	1713
CC	protein, and stimulating immune response in a patient. The present				
CC	sequence is one of the Human kidney tumour specific cDNAs. Note: The	QY	1039	TGTGCACAGTGTGAGGAGGACACCATCAAGTGGTGGCC-----GTGGACCGGAGC	1086
CC	sequence data for this patent did not form part of the printed				
CC	specification, but was obtained in electronic format directly from USPTO	Db	1714	TGTGTCGGCGTGCAAGCTGACACCTCGCGGTGGTCAATGACCGGATGCTGGGAGCCCAAC	1773
CC	at seqdata.uspto.gov/sequence.html?docID=20030109434.				
XX		QY	1087	TGTGACATCGGCAACTTTCGAGCTCTGTGGAGGAGCAGCGGT	1127
SQ	Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;				
		Db	1774	TGTGACACAGGAGGCTTTGAGCTCATCTATAGCAGCCCT	1814
	Query Match 26.0%; Score 421.4; DB 7; Length 2538;				
	Best Local Similarity 63.8%; Preq. No. 5.3e-80;				
	Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;				
QY	80 GGGAGGAGGCGGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA				139
Db	754 GGGATCTGACAGGCGCCCTCCCTATGTGTCAGGAGACAAGCAGCGCCCAAGAAACAGGAGAA				813

RESULT 12
ACN37881
ID ACN37881 standard; cDNA; 2553 BP.
XX

Human ORFX ORF2769 polynucleotide sequence SEQ ID NO:5537.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

W0200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US008621.

31-MAR-1999; 99US-0127607P.

02-APR-1999; 99US-0127636P.

05-APR-1999; 99US-0127728P.

30-MAR-2000; 2000US-00540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB43005.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease.

Claim 5; Page 4718-4720; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; vasotropic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

Sequence 2881 BP; 640 A; 812 C; 840 G; 589 T; 0 U; 0 Other;

Query Match 26.0%; Score 421.4; DB 3; Length 2881;

Best Local Similarity 63.8%; Pred. No. 5.5e-80;

Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;

QY 80 GGGAGGAGGCGGAGCAGGACCTGAGCAGCAGGCGAAGATGCTGAAATGCTGAGGCAA 139

DB 754 GGGATCGTGNACGGCCCTCCCTATGTGCGAGGAGACAGCAGGCCCAAGAAACAGGAGAAA 813

QY 140 AGCTCAG-GGGCTCCCGGGGAGCTTGTGGACATCCGCTGCAAGGTGTGCCAGGCCCTAC 198
DB 814 AACCCAGTGTGTGTGTCGCCAGAGTTTGTGGATGAAGCTCTGTGTGCGTGCAGGAGTAC 873
QY 199 CTGGGGCAGCTGAGAGCATAGAGCATCGACACGTCAGCAGATGCGTGGAGGACCTCACT 258
DB 874 CTTAGCACTTGGCCCACTGGACATCGACAGGACCTGGAGGCCCGCTGTACTCTACC 933
QY 259 GAGGCCAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTCGTTATGCGATGCTTTTC 318
DB 934 CCCGAGGGTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGGTCCACGAAGGGGCGAA 993
QY 319 ATCTCCAAATTCAGAAATTAATTTTTCGACGTCCAGGCTCTGCTGATAGATACAGTCT 378
DB 994 CTGAGCACCTCGACACTCAGGTCAGGCTCGAGGACATCTCTGACGAGCTCGAGGCC 1053
QY 379 GTGAACCCCTCAGACGAGCATTTGACGGCTCCGGCAACATCTGGATTATAAGCCCGCGGCC 438
DB 1054 GTGGTACCCAGATAGACATGGAGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCC 1113
QY 439 AAGTCCCGGGGCGGAGACATAGTGTGCATGAGACCGTGTGGAGGAGATCTCTGAGCTGGCA 498
DB 1114 AAGTCCCGTGGACGAGGCATCATGTGCATGAGCACCCCTGGAGGAGATGCTGAAGCTGGT 1173
QY 499 GCTGCAGACACCCCTCTTTCCAGGGACAAACAGTGGGTGTTCAGAAAGTACATCGAGACG 558
DB 1174 AACGGCAACCCCGTGGTGTGATGAAGAGCGGCAAGTGGGTGTGCAAGAAATATATTGAGCG 1233
QY 559 CGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGCGACTGG 618
DB 1234 CCCTCTCATCTTTGGCCAAAGTTTTCGACATCAGACAGTGGTTCCTGGTAACTGACTGG 1293
QY 619 AACCCCTGACCATCTGGTTCTACAAGGAGATTAATTTGCGGTTCTCAACTCAGCGCTTC 678
DB 1294 AACCCACTTACGTTGGTTCTACCGGACAGCTATATCCGCTTTTCCACGACGCCCTTC 1353
QY 679 TCCTTGACAGCTGACAGCGGCATCTCCTGTGCAACAAACGCGCTCCAGAAAGTACTGT 738
DB 1354 TCCTTGAAAGAACCTGGACAACTCAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACCTG 1413
QY 739 AAGAATGATGTGGGCGCGAGCCCTGCTGCGCGCACACAAACATGTGGACACGACACAGG 798
DB 1414 GAGAACTATGCCATCGGCATTCACCTGCTCCGCCAGACAAACATGTGTTAGCCAGAGG 1473
QY 799 TTCAGAGGTACTCTGAGCGGCCAGGGCGGTGGCGCGCTGTGGGGCAGCGTCATCTACCCG 858
DB 1474 TTCAGGCGCCACTGCGAGGAGTGGGTGCCCCAAATGCTTGGTCCACCATCATCGTGCCT 1533
QY 859 TCATGAAGAAGGCCATTCGCCACGCCCATGAAGGTGGCCCGAGCACACGCTGGAGCCTCGC 918
DB 1534 GGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGCG 1593
QY 919 AAGAACAGCTTTGAGCTCTACGGGGTGACTTCTGCTCTTGGGAGGACTTCAGGCCCTGG 978
DB 1594 AAGCCAGCTTTGAGCTCTATGGCGTGTGCTTCTGTTGGGGAGGACTTCCAGGCCCTGG 1653
QY 979 CTGATCGAGATCAATTCAGCCCCACCATGCACCCCGTCCACGCGCGTCCAGCGGCCAGCTG 1038
DB 1654 CTGATTTGAGATCAACGCCAGCCCCACGATGGCACCCCTCCACAGCAGTCACTGCCCGCTC 1713
QY 1039 TGTGCAAGGTGAGGAGGACACCATCAAGGTGGCC-----GTGGACCGGACG 1086
DB 1714 TGTGCTGGGTGCAAGCTGACACCTCGCGCTGTGTCATTGACCGGAGGCTGGACCGCAAC 1773
QY 1087 TGTGACATCGGCAACTTCGAGCTCCTGTGGAGGACGCGGT 1127
DB 1774 TGTGACACAGGAGCCTTTGAGCTCATCTATAAGCAGCCCGT 1814

RESULT 14

AA158606

ID AA158606 standard; cDNA; 3001 BP.

XX AAI58606;
AC
XX
XX
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 809.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00662191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX P-PSDB; ANM39450.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Claim 1; SEQ ID NO 809; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytosstatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
XX Sequence 3001 BP; 662 A; 854 C; 846 G; 639 T; 0 U; 0 Other;
SQ
Query Match 26.0%; Score 421.4; DB 4; Length 3001;
Best Local Similarity 63.8%; Pred. No. 5.5e-80;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 80 GGGAGAGCGCGGAGCAGCACCTTCAGCAGCAGGCGAAGATGCTGAATGCTGAGGCAA 139
DB 1223 GGGATCGTACAGGCCCTCCCTATGTGACAGAGACAGAGCCCAAGAAACAGGAGAAA 1282
QY 140 AGCTCAG-GGGGCTCCCGGGGCGAGCTTGTGGACATCGCGTGCAGAGGTGTGCCAGGCTAC 198

Db 1283 AACCCAGTGTGTGTGTCGCCAGAGTTTGTGGATGAAGCTCTGTGTGCCGAGGAGTAC 1342
QY 199 CTGGGGCAGCTGGAGCATGAGGACATCGACAGTTCAGCAGATGTCCTGGAGGACCTCACT 258
Db 1343 CTTAGCAACTTTGGCCCAATGACATCGACAAAGGACCTGGAGGCCCTGTGTACCTCACC 1402
QY 259 GAGGCGCAGTGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGCCGATGCTTTC 318
Db 1403 CCCGAGGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGGTCCACGAAGGGGAGAA 1462
QY 319 ATCTCCAAATTCAGAAATTAATTTTCGAGTGCAGGCTCTGTCTGAATGAATCAAGTCT 378
Db 1463 CTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCC 1522
QY 379 GTGAACCTTCAGACGACATTCAGCGGCTCCGGAACATCTGGATTTAAGCCCGCGGCC 438
Db 1523 GTGGTACCCCAATAGACATGGAAGGGGATCGCAACATCTGGATCTGGAAGCAGGAGCC 1582
QY 439 AAGTCCCGGGCGGAGACATAGTGTGATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGCA 498
Db 1583 AAGTCCCGGGAGGAGGATCATGTGATGGACCACTTGAGGAGATGCTGAAGCTGGTG 1642
QY 499 GGTGCAGACCACTCTTTCCAGGGGACAAAGTGGTGGTCCAGAAATGATCATCGAGACG 558
Db 1643 AACGGCAACCCCGTGTGTGATGAAGGACGGCAAGTGGTGGTGCAGAAATATATTGAGCGG 1702
QY 559 CGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCTCGTCAGGACTGG 618
Db 1703 CCCCTCTCATCTTTGGCACCAAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGACTGG 1762
QY 619 AACCCCTCAGCATCTGGTTCTACAGGAGAGTTACTTCGGGTTCTCAACTCAGCGCTTC 678
Db 1763 AACCCACTTACCGTGTGGTTCTACCGGACAGCTATATCGGCTTTTCCAGCAGCCCTTC 1822
QY 679 TCCCTGGACAAGCTGAGCAGCGCCATCCACTGTGCAACAAAGCGCTCCAGAAATGACTCG 738
Db 1823 TCCCTGAAGAACCTGGCAAACTCAGTGCACTGTGCAACAACTCCATCCAGAAAGCACCTG 1882
QY 739 AAGATGATGTGGCGCGCAGCGCCCTGCTCCCGCACACAAACATGTGGACGACCAAGG 798
Db 1883 GAGAACTCATGCGCATCGGCATCCACTGTCTCCGCCAGACAAACATGTGGTCTAGCCAGAG 1942
QY 799 TTTCCAGGAGTACTGTGAGCGCCAGCGCCGTGGCGCGCTGTGGGGCAGCGCTCATCTACCCG 858
Db 1943 TTTCCAGGCCCACTTCAGGAGATGGTGGCCCCCAATGCTTGGTCCACCATCATCTGTCCT 2002
QY 859 TCCATGAAGAAGGCCATCGCCCAACGCGCATGAAGGTGGGCCAGGACCAACGCTGGAGCCCTGC 918
Db 2003 GGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGACACCGGTGCAAGTGTCCG 2062
QY 919 AAGAACGCTTTGAGCTCTACGGGGCTGACTTCGTCTCTTGGGAGGAGACTTCAGGCCCTTG 978
Db 2063 AAGGCCAGCTTTGAGCTCTATGGCGTGTGCTTCTGGGGAGGACTTCCAGCCCTTG 2122
QY 979 CTGATCGAGATCAATTCAGAGCCCCCATGTCACCCCGTCCAGCCCGGTCAAGGCCAGCTG 1038
Db 2123 CTGATTGAGATCAACGCCAGCCCCCAGTGGCAACCTTCCAGCAGTCACTGCGCCGCTC 2182
QY 1039 TGTGCACAGGTGCAGAGGACACCATCAAGGTGGCC-----GTGACCCGAGC 1086
Db 2183 TGTGCTGGCGTGCAAGCTGACACCCCTCGCGTGGTCAATTGACCGGATGTGGACCGCAAC 2242
QY 1087 TGTGATCGGCAACTTTCAGGCTCTGTGGAGGAGCCCGGT 1127
Db 2243 TGTGACACAGGAGCCTTTGAGCTCATCTATAAGACGACCCGT 2283
RESULT 15
ADQ98824
ID ADQ98824 standard; cDNA; 3001 BP.
XX
AC ADQ98824;

XX 23-SEP-2004 (first entry)
XX DNA encoding human GPCR-like protein seqid 494.
XX
XX ophthalmological; immunomodulatory; cytostatic; antiatherosclerotic;
KW antidiabetic; GPCR-like protein; ophthalmic disorder;
KW neurological disorder; immunological disorder; nephritic disorder;
KW hormonal dysfunction; cancer; atherosclerosis; diabetes;
KW molecular weight marker; food supplement; human; ss.
XX Homo sapiens.
XX US6569662-B1.
XX 27-MAY-2003.
XX
XX 19-JUL-2000; 2000US-00620312.
XX
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Zhou P, Drmanac RT;
XX WPI; 2001-442255/47.
XX
XX New G-protein-coupled receptor-like polypeptides and polynucleotides,
PT useful for treating diseases of ophthalmic, neurological, immunological
PT and nephritic systems and hormonal dysfunction, cancer, atherosclerosis
PT and diabetes.
XX
XX Example 2; SEQ ID NO 494; 92pp; English.
XX
XX The invention describes an isolated polynucleotide (I) comprising a fully
CC defined (S1) of 749, 3188, 2484, 1169, 2936, 1467, 5773, 5714, 4041,
CC 1372, 3996, 3945, 2735, 1788, 585, 1782, 327, 5714 or 2282 nucleotides as
CC given in the specification, its translated or protein coding portion, its
CC extracellular portion or its active domain. The GPCR-like polypeptides
CC and polynucleotides are useful for the treatment of diseases of
CC ophthalmic, neurological, immunological and nephritic systems. They may
CC also be used to treat hormonal dysfunction, cancer, atherosclerosis and
CC diabetes. The antibodies are useful for detecting or quantitating the
CC polypeptide in tissue. The polypeptides can also be used as molecular
CC weight markers and as a food supplement. This sequence represents a human
CC polynucleotide of the invention.
XX
XX Sequence 3001 BP; 562 A; 854 C; 846 G; 639 T; 0 U; 0 Other;
SQ
Query Match 26.0%; Score 421.4; DB 5; Length 3001;
Best Local Similarity 63.8%; Pred. No. 5.5e-80;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 80 GGGAGGAGGCGGAGCAGCAGCTGAGCAGCAGGCAAGAGTGTGAAATGCTGAGGCA 139
DB 1223 GGGATGTGACAGGCCCTCCCTATGTGACAGGAGCAAGCAGGCCCAAGAAACAGGAGAA 1282
QY 140 AGCTCAG-GGGCCCTCCGGGCGAGCTTGTGACATCGCGTCAAGGTGTGCCAGGCTAC 198
DB 1283 AACCCAGTGTGTGTCCTCCAGAGTTGTGATGAAGCTGTGTGCGTGAGAGGTAC 1342
QY 199 CTGGGCGAGCTGGAGCATGAGGACATGACACATGTCAGAGATCCGCTGGAGGACCTCACT 258
DB 1343 CTTAGCAACTTGGCCACATGAGCATGACAGGACCTGGAGGCCCGCTGTACCTCACC 1402
QY 259 GAGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTATGCGGATGCTTC 318
DB 1403 CCCGAGGGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGTGTCCAGAGGGGAGAA 1462
QY 319 ATCTCCAATTCAGAAATTAATTTTCGAGTGCAGGCTCTGCTGAATAGATCACTGCT 378
DB 1463 CTCAGGCACTCGACACTAGGTCCAGCGCTGTGAGGACATCTCTGCGAGGTGCGAGGCC 1522

QY 379 GTGAACCTCTCAGACGGACATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGGCC 438
DB 1523 GTGGTATCCCCAGATAGACATGGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCC 1582
QY 439 AAGTCCCGGGGCGAGACATAGTGTGATGCGACCGTGTGGAGGAGATCTCTGGAGCTGGCA 498
DB 1583 AAGTCCCGGGGAGGAGCATCATGTGATGAGACCACTCTGGAGGAGATGCTGAAGCTGGT 1642
QY 499 GCTGCAAGACACCTCTTTCCAGGGAACAAGTGGGTGTCAGAAAGTATCATCGAGACG 558
DB 1643 AACGGCAACCCCGTGTGTATGAAGGACGGCAAGTGGGTGTCAGAAAGTATATTGAGCGG 1702
QY 559 CCGCTGCTCATCTGTGACACCAAGTTGACATCAGACAGTGGTTCCTCTCAGCGACTGG 618
DB 1703 CCCCTCTCATCTTTGGCACCAAGTTGACCTCAGACAGTGGTTCCTCTGGAAGTGGT 1762
QY 619 AACCCCTGACCATCTGGTTCTCAAGAGGAGTACTTTGGCGTTCTCAACTCAGCGCTTC 678
DB 1763 AACCCACTTACCGTGTGTTCTACCGGACAGTATATCGCTTTTCCACGCGAGCCCTTC 1822
QY 679 TCCTTGGAACAAGCTGGAAGCGCCATCCACTGTGCAACAAACCGCTCCAGAAAGTACTG 738
DB 1823 TCCTTGAAGAACCTGGACAACTCAGTGCACCTGTGCAACAACTCCATCCAGAAAGCCTG 1882
QY 739 AGAATGATGTGGGCGGACAGCCCTGCTGCCCGGACACAAATGTGGACCAAGCAGG 798
DB 1883 GAGAATCATGCGCATCGGCATCCACTGCTTCCGCCAGACAAATGTGGTCTAGCCAGAGG 1942
QY 799 TTCACGAGTACTGTGACGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGTCACTTACCCG 858
DB 1943 TTCACGCGCCACTGTGACGAGATGGGTGCCCAATGCTTGGTCCACCATCATGTGCTC 2002
QY 859 TCCATGAAGAAGGCCATCGCCCAATGAAGGTGGCCCGAGGACCAAGTGGAGCTTCG 918
DB 2003 GGCAATGAAGGATGCTGTGATCCAGCGCACTTCAGACCTCCAGGACACCGTGCAGTGC 2062
QY 919 AGAAGAGCTTTGAGCTCTACGGGGCTGACTTGTCTCTTGGGAGGAGACTTCAGGCGCT 978
DB 2063 AAGGCCAGCTTTGAGCTCTATGGCGCTGACTTGTGTTCGGGGAGGACTTCCAGCGCT 2122
QY 979 CTGATCGAGATCAATTCACAGCCCAACATGCAACCGCTCCACGCGGTTCAGGCGCCAGCT 1038
DB 2123 CTGATGAGATCAACCCAGCCCGCCAGTGGCAACCTCCACAGCAGTCACTGCGCGCTC 2182
QY 1039 TGTGCAAGTGTGAGGAGGACACCATCAAGGTGGCC-----GTGGACCGCAGC 1086
DB 2183 TGTGCTGGCGTGCAAGCTGACACCTCGCGTGGTCAATTGACCGGATGCTGGACCGCAAC 2242
QY 1087 TGTGACATCGCAACTTTCAGCTCTCTGTTGGAGGCGCCGGT 1127
DB 2243 TGTGACACAGGAGCCTTTGAGCTCATCTATAAGCAGCCCGT 2283

Search completed: September 24, 2005, 18:03:27

Job time : 987.756 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 24, 2005, 11:30:53 ; Search time 7533.96 Seconds
(without alignments)
10438.446 Million cell updates/sec

Title: US-10-615-659-1_COPY_153_1775
Perfect score: 1623
Sequence: 1 atggcattcagcattcctca.....cgccggagaaagggttca 1623

Scoring table: IDENTITY NUC
Gapex 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sv.*
- 13: gb.un.*
- 14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	817	50.3	817	6	CQ724907 Sequence
2	501	30.9	101270	9	HS355C18
3	432.8	26.7	2380	6	AX301197 Sequence
4	431.6	26.6	2326	6	BD158727 Primer fo
5	431.6	26.6	2326	6	AX881015 Sequence
6	431.6	26.6	2326	9	AK023960 Homo sapi
7	431.6	26.6	4238	9	HSMB05098
8	421.4	26.0	2553	10	AR39003 Sequence
9	421.4	26.0	3001	6	AR339003
10	418	25.8	2848	6	AX834642
11	418	25.8	2848	9	AK097236 Homo sapi
12	380.2	23.4	1684	9	AF078842
13	366.4	22.6	1857	10	BC006830
14	283.6	17.5	163783	10	AC119959
15	283.6	17.5	201420	2	AC117700
16	277.2	17.1	205949	2	AC134940
17	277.2	17.1	218249	2	AC097425
18	231.4	14.3	5282	6	CQ842940
19	231.4	14.3	5282	9	AK125875

C	20	224.6	13.8	3828	6	CQ850619	Sequence
C	21	224.6	13.8	3828	9	AK127786	Homo sapi
C	22	213	13.1	2979	6	CQ595968	Sequence
C	23	192.2	11.8	145435	2	AC026685	Homo sapi
C	24	192.2	11.8	161903	2	AC021996	Homo sapi
C	25	191.4	11.8	1560	9	BC009479	Homo sapi
C	26	190.6	11.7	1958	6	BD160681	Primer fo
C	27	190.6	11.7	1958	6	AX884044	Sequence
C	28	190.6	11.7	1958	9	AK024110	Homo sapi
C	29	190.6	11.7	155313	2	AC068315	Homo sapi
C	30	190.6	11.7	185067	9	AC022382	Homo sapi
C	31	190.6	11.7	189430	2	AC011610	Homo sapi
C	32	190.6	11.7	191834	2	AC026196	Homo sapi
C	33	190.6	11.7	197360	2	AC018829	Homo sapi
C	34	190.6	11.7	216511	9	AC018809	Homo sapi
C	35	181.2	11.2	224077	2	AC120010	Homo sapi
C	36	181.2	11.2	266888	2	AC137881	Mus muscu
C	37	174.2	10.7	2543	6	CQ595656	Sequence
C	38	146.4	9.0	220290	2	BX936393	Danio rer
C	39	145.8	9.0	161903	2	AC021996	Homo sapi
C	40	145.8	9.0	222594	2	AC016947	Homo sapi
C	41	144.8	8.9	4615	6	CQ595655	Sequence
C	42	144.8	8.9	76854	3	AC003052	Drosophil
C	43	144.8	8.9	86398	2	AC017785	Drosophil
C	44	144.8	8.9	167201	3	AC092397	Drosophil
C	45	144.8	8.9	259718	3	AE003614	Drosophil

ALIGNMENTS

RESULT 1	CQ724907	817 bp	DNA	linear	PAT 03-FEB-2004
LOCUS	Sequence 10841 from Patent WO02068579.				
DEFINITION	CQ724907				
ACCESSION	CQ724907.1				
VERSION	GI:42285764				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	1				
AUTHORS	Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.				
TITLE	Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof				
JOURNAL	Patent: WO 02068579-A 10841 06-SEP-2002;				
FEATURES	PE Corporation (NY) (US)				
source	Location/Qualifiers				
	1..817				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				

ORIGIN	Query Match	50.3%;	Score 817;	DB 6;	Length 817;
	Best Local Similarity	100.0%;	Pred. No. 2.5e-129;		
	Matches 817;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	306	TGGCGATGTTTCAATCTCAATTCAGAAATTTACTTTTCGCAGTGCAGGCTCTGCTGAA	365		
Db	1	TGGCGATGTTTCAATCTCAATTCAGAAATTTACTTTTCGCAGTGCAGGCTCTGCTGAA	60		
Qy	366	TAGAATCAGTCTGTGAACCTCAGACGACATTGACGGCTCCGGAACATCTGGATTAT	425		
Db	61	TAGAATCAGTCTGTGAACCTCAGACGACATTGACGGCTCCGGAACATCTGGATTAT	120		
Qy	426	AAAGCCCGCGCAAGTCCCGGGCGGACATAGTGTGCATGGACCGTGTGGAGGAGAT	485		
Db	121	AAAGCCCGCGCAAGTCCCGGGCGGACATAGTGTGCATGGACCGTGTGGAGGAGAT	180		
Qy	486	CCTGGAGCTGGAGCTGCAGACACCTCTTTCCAGGGACACACAGTGGGTGCTCCAGAA	545		

Db	181	CTTGAGCTGGACGCTGAGACACCCTCTTTCCAGGGAACAAGTGGGTGGTCCAGAA	240
Qy	546	GTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTGCACATCAGACAGTGTTCCT	605
Db	241	GTAATCGAGACGCGCTGCTCATCTGTGACACCAAGTTGCACATCAGACAGTGTTCCT	300
Qy	506	CGTCACGACTGGAAACCCCTGACCATCTGGTTCCTACAAGGAGAGTTACTTTCGGTTC	665
Db	301	CGTCACGACTGGAAACCCCTGACCATCTGGTTCCTACAAGGAGAGTTACTTTCGGTTC	360
Qy	666	AATCTACGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACTGTGCAACAACGCCGT	725
Db	361	AATCTACGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACTGTGCAACAACGCCGT	420
Qy	726	CCAGAAGTACTGAGAAATGATGTGGCGCGGAGCCCTGCTGTCGCCCGCACACAACATGTG	785
Db	421	CCAGAAGTACTGAGAAATGATGTGGCGCGGAGCCCTGCTGTCGCCCGCACACAACATGTG	480
Qy	786	GACCAGACACAGGTTCCACGAGTAGTACTTCGACGCGCAGGGCCGCTGGCGCCCGTGGGGCAG	845
Db	481	GACCAGACACAGGTTCCACGAGTAGTACTTCGACGCGCAGGGCCGCTGGCGCCCGTGGGGCAG	540
Qy	846	CGTCATCTACCGTCCATGTAAGAAAGGCCATCGCCCAAGCATGAAGTGGCCAGGACCA	905
Db	541	CGTCATCTACCGTCCATGTAAGAAAGGCCATCGCCCAAGCATGAAGTGGCCAGGACCA	600
Qy	906	CGTGGAGCCTCGCAAGAAACAGCTTTGAGCTCTACGGGGCTGACTTCCTCTCTGGGAGGGA	965
Db	601	CGTGGAGCCTCGCAAGAAACAGCTTTGAGCTCTACGGGGCTGACTTCCTCTCTGGGAGGGA	660
Qy	966	CTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCAACCATGCAACCCGTCACGCGCGT	1025
Db	661	CTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCAACCATGCAACCCGTCACGCGCGT	720
Qy	1026	CAGGCCCAAGCTGTGTCACAGGTGCAAGGAGGACACCATCAAGGTGCGCTGGACCGCAG	1085
Db	721	CAGGCCCAAGCTGTGTCACAGGTGCAAGGAGGACACCATCAAGGTGCGCTGGACCGCAG	780
Qy	1086	CTGTGACATCGGCAACTTCGAGTCTCTGTGGAGGCAG	1122
Db	781	CTGTGACATCGGCAACTTCGAGTCTCTGTGGAGGCAG	817

RESULT 2	
HS355C18/c	
LOCUS	HS355C18 101270 bp DNA linear PRI 05-JUN-2003
DEFINITION	Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3 Contains the KIAA0027 gene, ESTs, STSS, GSSs and seven putative CpG islands, complete sequence.
ACCESSION	AL022327
VERSION	AL022327.17 GI:5304851
KEYWORDS	HTG; CpG island; KIAA0027.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 101270)
TITLE	Cobley, V.
JOURNAL	Direct Submission
COMMENT	Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humbug@sanger.ac.uk On Jun 30, 1999 this sequence version replaced gi:5262834. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Enr.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information

```

on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RP3-355C18 is from the library RPCI-3 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCIPAC2
This sequence is the entire insert of clone RP3-355C18 The true
left end of clone RP5-898I4 is at 5390 in this sequence.

FEATURES
      source
      1..101270
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="22"
        /map="q13.3"
        /clone="RP3-355C18"
        /clone_lib="RPCI-3"
      1..1038
        /note="match: STS: Em:AL022483"
        /complement(1..242)
        /note="AluJo repeat: matches 1..237 of consensus"
      repeat_region
      338..484
        /note="L1M4 repeat: matches 829..959 of consensus"
      repeat_region
      485..792
        /note="AluSx repeat: matches 2..305 of consensus"
      repeat_region
      795..1095
        /note="AluSx repeat: matches 1..299 of consensus"
      repeat_region
      1096..1165
        /note="L1M4 repeat: matches 960..1031 of consensus"
      repeat_region
      1166..1563
        /note="MSTB1 repeat: matches 4..432 of consensus"
      repeat_region
      1564..1750
        /note="L1M4 repeat: matches 1031..1203 of consensus"
      repeat_region
      1757..2038
        /note="AluSx repeat: matches 1..282 of consensus"
      repeat_region
      complement(2054..3342)
        /note="SVA repeat: matches 1..1386 of consensus"
      repeat_region
      3343..3359
        /note="2.8 copies 6 mer TCTCTT 34% conserved"
      repeat_region
      3360..3417
        /note="L1R39-int repeat: matches 3904..3961 of consensus"
      repeat_region
      3418..3577
        /note="L1M4 repeat: matches 1203..1363 of consensus"
      repeat_region
      3578..3861
        /note="AluJo repeat: matches 1..300 of consensus"
      repeat_region
      3862..4334
        /note="L1M4 repeat: matches 1363..1830 of consensus"
      repeat_region
      4341..4489
        /note="L1M4 repeat: matches 2065..2311 of consensus"
      repeat_region
      complement(4492..4840)
        /note="MER4A1 repeat: matches 96..472 of consensus"
      repeat_region
      4990..4999
        /note="2.5 copies 4 mer CTGT 20% conserved"
      repeat_region
      complement(5142..5300)
        /note="AluJb repeat: matches 137..295 of consensus"

```


ACCESSION	AX3011197
VERSION	AX3011197.1
KEYWORDS	GI:17382288
SOURCE	.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	Yue,H., Tang,Y.T., Au-Yang,J., Lu,D.A., Baughn,M.R., Hillman,J.L., Azimzai,Y., Lal,P., Yao,M.G., Bandman,O., Burford,N., Batra,S., Kearney,L. and Policky,J.L.
JOURNAL	Cytoskeleton-associated proteins
FEATURES	Patent: WO 0185942-A 39 15-NOV-2001;
source	Incyte Genomics, Inc. (US) Location/Qualifiers 1..2380 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606" /note="Incyte ID No: 2156553CB1"
ORIGIN	
Query Match	26.7%; Score 432.8; DB 6; Length 2380;
Best Local Similarity	59.1%; Pred. No. 5.8e-64;
Matches	802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;
Qy	108 CAGCAGGCAAGATGCTCAAATGCTGAGGCCAAAGCTCAG-GGGCCCTCCCGGGCAGCTTG 166
Dd	361 CAGGAGACAGCAGCCCAGNAACAGAGAAAACCAGGTGTGGTGTCCCCAGAGTTG 420
Qy	167 TGGACATCGCGTCGAAGGTGTGCAGGCCCTACTCTGGGCGAGCTGGAGCATGAGGACATCG 226
Dd	421 TGGATGAAGCTCTGTGTGCGTGCAGGAGTACCTTAGCAACTTGGCCCAACATGGACATCG 480
Qy	227 ACAGCTCAGCAGATCGCGTGGAGACCTCACTGAGCGCGAGTGGAGGACCTGACCCAGC 286
Dd	481 ACAAGGACCTGGAGGCCCGCGTGTACCTCACCCCGAGGGGTGGTCTCTCTCTCCAGC 540
Qy	287 AGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAATTACTTTTCGC 346
Dd	541 GCTACTACCAAGTGTCCAGAAGGGCGAACACTCAGGCACTTCGACACTCAGGTCGAGC 600
Qy	347 AGTGCCAGGCTCTGCTGAATAGAAATCACTGTGTGAACCCCTCAGACGGACATTCAGCGGC 406
Dd	601 GCTGTGAGGACATCTCTCAGCAGCTGCAGCGCGTGGTACCCAGATAGACATGGAAGGG 660
Qy	407 TCGGGAACATCTGGATTATAAGCCCGCGGCAAGTCCCGGGGCGGAGACATAGTGTGCA 466
Dd	661 ATCGCAACATCTGGATCGTGAAGCCAGGAGCAGAGTCCCGTGGACGAGGATCATGTGCA 720
Qy	467 TGACCGTGTGGAGGAGATCTCGAGCTGCGAGCTGCAGCTGCAGACCACTCTTTCAGAGGACA 526
Dd	721 TGAACCACTTGGAGGAGATGCTGAGCTGGTGAACGGCAACCCCGTGGTGAATGAAGGACG 780
Qy	527 ACAAGTGGTGGTCCAGAAGTACATCAGAGACGCGCGTCTCATCTGTGACACCAAGTTTCG 586
Dd	781 GCAAGTGGTGGTCCAGAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAAGTTTG 840
Qy	587 ACATCAGACAGTGGTTCTCTGTCACGACGTGGAAACCCCTTGACCACTGTGTTCTTACAAGG 646
Dd	841 ACCTCAGACAGTGGTTCTCTGTAACTGACTGGAACCCCACTTACCGTGTGGTTCTTACCGCG 900
Qy	647 AGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACGCGCCATCC 706
Dd	901 ACAGCTATATCCCGTTTTTCCACGCGAGCCCTTCTCCCTGAAAGAACCTTGGACAACTCAGTGC 960
Qy	707 ACCTGTGCACAAACCGCGTCCAGAAGTACCTGAAGAATGATGTGGGCGCGAGCCGCCCTTCG 766
Dd	961 ACCTGTGCACAACTCCATCCAGNAGACCTGGAGAACTCATGCCATCGGCATCCACTGC 1020
Qy	767 TGCCCGCACACAACTGTGGACCAACACAGGTTTCAGGAGTACCTGCAGCGCCAGGCCC 826
Dd	1021 TTCGCGCAGACAACTGTGTCTTAGCCAGAGGTTTCAGGCGCCACTTCAGAGATGCGGT 1080

	10,	PC	C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12NS/00 CC	Primer for synthesizing full-length cDNA and use thereof PH Key
FEATURES	FT	CDS	Location/Qualifiers (94)..(2325).	
SOURCE			Location/Qualifiers 1..2326 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"	
ORIGIN				
	Query Match	25.6%;	Score 431.6; DB 6; Length 2326;	
	Best Local Similarity	63.5%;	Pred. No. 9.3e-64;	
	Matches 698; Conservative	0;	Mismatches 389; Indels 13; Gaps 2;	
Qy	108	CAGCAGGCAAGATCGTGAATAATGCTGAGGCCAAAGCTCAG-GGGCTCCCGGGGCAGCTTG 166		
Db	620	CAGAGACAGCACCCAGAAACAGAGAAAACCAGTGTGGTGTCCCAGAGTTTG 679		
Qy	167	TGGACATCGCGTGCAAGGTGTGCCAGGCCTACCTGGGGCAGCTGGAGCATGAGGACATCG 226		
Db	680	TGGATGAAGCTCTGTGTGCGTCGCGAGGTACTCTTAGCACTTGGCCCCATGACATCG 739		
Qy	227	ACAGCTCAGCAGATGCCGTGTAGAGACCTCACTGAGGCCGAGTGGAGGAGACCTGACCCAGC 286		
Db	740	ACAAGGACCTGGAGGCCCGGTGTACCTACCCCGAGGGGTGGTCTCTTCTCTCCAGC 799		
Qy	287	AGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCATTCAAGAAATTACTTTTCGC 346		
Db	800	GCTACTACCAAGTGGTCCAGAAGGGGCAAGTAACCTCAGCACCTCGACACTCAGTTCAGC 859		
Qy	347	AGTGCCAGGCTCTGCTGAATAGAATACGTCGTGTGAACCCCTCAGACGGAATTCACGGGC 406		
Db	860	GCTGTGAGACATCTCGACGAGCTGCAGGCCGTGTGTACCCAGATAGACATGGAAGGGG 919		
Qy	407	TCGGAACATCTGGATTATTAAGCCCGCGCAAAGTCCCGGGGCCGAGACATAGTGTGCA 466		
Db	920	ATCGCAACATCTGATCGTGAAGCAGGAGCAAAGTCCCGCGGACGAGGCATCATGTGCA 979		
Qy	467	TGGACGTGTGGAGGAGATCTGAGCTGGCAGCTGCAGACCAACCCCTCTTCCAGGACA 526		
Db	980	TGGACCACTGGAGGAGATGCTGAAGCTGTGTGAACGGCAACCCCGTGTGATGAAGGACG 1039		
Qy	527	ACAAGTGGGTGGTCCAGAAGTACATCGAGAGCCGCTGCTCATCTGTGACACCAAGTTTCG 586		
Db	1040	GCAAGTGGGTGGTGCAGAAATATATTGAGCGGCCCTCTCATCTTTGGACACCAAGTTTG 1099		
Qy	587	ACATCAGCAGTGTGTTCTCGTCACGAGCTGGAACCCCTTGACCATCTGTGTTCTACAAGG 646		
Db	1100	ACCTCAGACAGTGGTTCTCGTTAATCTGACTGGAACCCCACTTACCGTGTGGTTCTACCGCG 1159		
Qy	647	AGAGTTACTTGGGTTTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCCCATCC 706		
Db	1160	ACAGCTATATCCGCTTTTCCAACGAGACCCCTTCTCCCTGAAGAACCTGGACAACTCAGTGC 1219		
Qy	707	ACCTGTGCAACACGCGCTGCAGAAATACCTGAAGAATGATGTGGCGCGCAGCCCTCTGC 766		
Db	1220	ACCTGTGCAACACTCCATCCAGAAGACACTTGGAGAACTCATGCCATCGGCATCCACTGC 1279		
Qy	767	TGCCCGCACACAAATGTGGACACGACACAGGTTCCAGGAGTACTCTGCAGCGCCAGGGCC 826		
Db	1280	TTCCGCCAGACAAATGTGTGCTAGCCAGAGGTTCCAGGCCCACTTCAGGAGATGGTG 1339		
Qy	827	GTGGCGCGGTGTGGGGAGCGTCACTTACCGGTCCATGAAGAAGGCCCATCGCCCAACGCCA 886		
Db	1340	CCCCAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGATGCTGTGATCCACGCAC 1399		
Qy	887	TGAGGTGGCCACGAGCACGTGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTG 946		
Db	1400	TTCAGACTCCACGAGACCGGTGCAGTGTGGGAAGGCCAGCTTTTGTAGCTCTATGGCGCTG 1459		
Qy	947	ACTTTCGCTTTGGGAGGAGCTTCAGGCGCTCGGCTGATCGAGATCAATTTTCAGGCCCAACCA 1006		

Db 680 TGGATGAAGCTCTGTGTGTCGTCGAGGAGTACCTTAGCAACCTTTGGCCCCATGGACATCG 739
QY 227 ACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGCGCCAGTGGGAGGACCTTGACCCAGC 286
Db 740 ACNAGACCTGGAGGCCCGCTGTACTTACCCCCGAGGGCTGGTCCCTTCTTCCCTCCAGC 799
QY 287 AGTACTACTCCCTCGTTCATGCGGATGCTTTCATCTCCAAATTCAGAAATTAATTTTCGC 346
Db 800 GCTACTACCAAGTGGTCCACGAAGGGGCAGAACTCAGGCACCTCCGACACTCAGGTCACGC 859
QY 347 AGTCCAGGCTCTGCTGTAATAGATCACCTCTGTGAACCTTCAGCGGACATTTGACGGC 406
Db 860 GCTGTGAGGACATCTCTGACAGCTCGACGCCCTGTGTATCCACAGATAGCATGGAAAGGG 919
QY 407 TCCGGAACATCTCGATTATAAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGCA 466
Db 920 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAGTCCCGCGGACGAGGCATCATGTGCA 979
QY 467 TGGACCGTGTGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACA 526
Db 980 TGGACCACTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGTGTGATGAAGGACG 1039
QY 527 ACAAGTGGTGTCCAGAGTACATCGACCGCGCTGCTCATCTGTGACACCAAGTTTCG 586
Db 1040 GCAAGTGGTGTGTGCAAGATATATTGAGCGGCCCTCTCTCATCTTTGGCACAAGTTTG 1099
QY 587 ACATCAGACAGTGGTTCCTCGTCACGACTGGAACCCCTGACCACTCTGGTTCACAAAG 646
Db 1100 ACTTCAGACAGTGGTTCCTGTTAACTGACTGNAACCACTTACCGTGTGTTCTACCGCG 1159
QY 647 AGAGTTACTTGGCGTTCCTCAACTAGCGGTTCTTCCCTGCAAGCTTGGACAGCGCATCC 706
Db 1160 ACAGCTATATCGGCTTTTCCAGCGCAGCCCTTCTCCCTGAAGAACCTGGCAAACTCAGTGC 1219
QY 707 ACTGTGCAACACGCGTCCAGAGTACTGGAAGATGATGTGGCGCCGACGCCCTCTGC 766
Db 1220 ACTGTGCAACAACTCCATCCAGAAGCACTGGAGAACTCATGCGCATCGGCATCCACTGC 1279
QY 767 TGCCTCGCACACAACATGTGGACAGCAGCAGGTTTCAGGAGTACCTGACGCGCCAGGGCC 826
Db 1280 TTCCGCCACACAACATGTGTTAGCCAGAGGTTCCAGGCCCACTGACGAGATGGGTG 1339
QY 827 GTGGCGCCGTGTGGGCGACGCTCATCTACCGTCCATGAAGAAGGCGCATGCGCCACGCCA 886
Db 1340 CCCCAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGGATGCTGTGATCCACGCAC 1399
QY 887 TGAAGTGGCCCGGACACACGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTG 946
Db 1400 TTCAGACCTCCACGAGACCGGTGAGTGTGCGAAGCCAGCTTTGAGCTCTATGGCGCTG 1459
QY 947 ACTTCGTCTTGGGAGGACTTTCAGGCCCTGGTGTATCGAGATCAATTCAGGCCCCACCA 1006
Db 1460 ACTTCGTGTTCGGGAGGACTTTCAGCCCTGGCTGATTGAGATCAACGCCAGCCCCACGA 1519
QY 1007 TGCACCCGTCCACGCGGTTCAGGCCCGCCAGCTGTGTGCAAGGTGACGAGGAGCACCATCA 1066
Db 1520 TGGCACCTCCACAGCAGTCACTGCGCCGCGCTCTGTGCTGGCGTGCAGAGTGAACCCCTGC 1579
QY 1067 AGGTGGCC-----GTGGACCGAGCTGTGACATCGCACTTCGAGCTCTCTGT 1114
Db 1580 GCGTGGTTCATTACCGGATGCTGGACCGCAACTGTGTGACACAGGAGCTTTTGAGCTCATCT 1639
QY 1115 GGAGGCGAGCGGTGGTTGAGCGCGCCCGCCCAATTCAGCGGGTCCCACTCTCGGTGGCGGGG 1174
Db 1640 ATAGCAGCGCTCTGTGGAGGTGCCTCATATGTGGGCATCGGCTCTCTGTAGAGGGCT 1699
QY 1175 TCAGTGTGAGGAGACGAGC 1194
Db 1700 TCACCATCAAGAAGCCCATG 1719

RESULT 6
AK023960

LOCUS AK023960 2326 bp mRNA linear PRI 30-JAN-2004
DEFINITION Homo sapiens cDNA FLJ13898 fis, clone THYRO1001738, weakly similar to TUBULIN--TYROSINE LIGASE (EC 6.3.2.25).
ACCESSION AK023960
VERSION AK023960.1 GI:10436083
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Chordata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Negai,K., Kimura,K., Makita,H., Sekine,M., Oyabashi,M., Nishi,T., Shibahara,T., Tanaka,T., Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y., Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M., Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H., Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yosida,M., Horuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Oshima,A., Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Ichihara,T., Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T., Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shinizu,P., Wakebe,H., Hishigaki,H., Watanabe,T., Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K., Kumagai,A., Tanigami,A., Fujimori,Y., Fujimori,Y., Komiyama,M., Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hiroa,M., Ohmori,Y., Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T., Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J., Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K., Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R., Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
Complete sequencing and characterization of 21,243 full-length human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
2
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2326)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
Location/Qualifiers
1..2326
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYRO1001738"
/tissue_type="thyroid gland"
/clone_lib="THYRO1"


```
/db_xref="GI:21739507"
/db_xref="GOA:Q8NDN8"
/translation="WDIDKDLAPLYLTPGKSLFLQRYVYVHGAELRLDQVOR
CEDILQQLQAVVPOIDMEGRNIWIYKPKAKSRGRGIMCMHLEMLKLIVNGPVVMK
DKRWVQKYLIERPLLIIFGKFDLRQWFLVDFYRDSYIRFSTQPSFLKNLD
NSVHLNNSIQKHELENSCHRPFLPDNNMSSORFOAHLQEMGANWSTIIVPGMKD
AVIHALQTSQDTVQCRKASFLYADVFEGEDFQPLWIEINASPMPASTAVTARICA
GVOADLRVVIDRRLDRNCDTGAPELIYKQAPVEVPOVYVIRLLVEGFTIKKPMAMCH
RMGVPRVAPVLLTQBGSGEARHHPPSLHTKAQLPSPHVLRHOGQVLRHOSKLVGTKA
LSTTGKALRILFTAKVVISLPPNLDPKVAPSLIKPRKVGDL"
4216
polyA_site
/gene="DKFZp586B0320"

ORIGIN
Query Match          26.6%; Score 431.6; DB 9; Length 4238;
Best Local Similarity 63.5%; Pred. No. 8.4e-64;
Matches 698; Conservative 0; Mismatches 389; Indels 13; Gaps 2;

QY 108 CAGCAGCAAGATGCTGAAAATGCTCAGGCAAGCTCAG-GGGCCCTCCCGGGGCGAGCTTG 166
DB 1430 CAGGAGCAAGCAGCCCAAGAACACAGAGAAAACCCAGTGTGGTGTCCCCAGATTG 1489
QY 167 TGGACATCGCGTCAAGGTGTGTCAGGCCTACCTGGGGCAGCTGGAGCATGAGGACATCG 226
DB 1490 TGGATGAAGCTCTGTGTGCGTGCAGAGAGTACCTTAGCAACTTGGGCCCATGGAATCG 1549
QY 227 ACACGTCAGCAGATCCGCTGGAGGACCTCACTGAGCCGAGTGGGAGGACCTGACCCAGC 286
DB 1550 ACAGGACCTGGAGGCCCGCTGTACTCACCCCGGAGGCGTGTCCCTCTTCCTCCAGC 1609
QY 287 AGTACTACTCCCTCGTTTCATGCGGATGCTTTCATCTCCAAATCAAGAAATTACTTTTCGC 346
DB 1610 GCTACTACCAAGTGTCTCAGAGGGCGAGAACTCAGGACCTCGACACTCAGGTCAGC 1669
QY 347 AGTGCCAGCGCTGTGTAATAGAAATCAGCTGTGTGTAACCCCTCAGACGGAATGACGGGC 406
DB 1670 GCTGTGAGGACATCTCTGCAGCAGCTGCAGCGCGTGTGTAACCCAGATAGAGGAGGG 1729
QY 407 TCGGAACATCTGGATTATTAAGCCCGCCGCAAGTCCCGGGCGCAGACATAGTGCA 466
DB 1730 ATCGCAACATCTGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGGACATGATGCA 1789
QY 467 TGGACCGCTGTGAGGAGATCTCTGAGCTGGCAGCTGCAGACCAACCCCTCTTTTCAGGGACA 526
DB 1790 TGGACCACCTGGAGGAGATGCTGAAGCTGTGTGAACCGCAACCCCGTGTGATGAAGGACG 1849
QY 527 ACAAGTGGGTGTGCAAGAGTACATCAGACGCGCTGTCTCATCTGTGTGACCAAGATTCCG 586
DB 1850 GCAAGTGGGTGTGCAAGAGTATATTGAGCGCGCCCTCTCTCATCTTTTGGCACCAGTTTG 1909
QY 587 ACATCAGACAGTGGTTCCTCGTCACGAGCTGGAAACCCCTGACCATCTCGTTCTTCAAGG 646
DB 1910 ACCTCAGACAGTGGTTCCTCGTGAACCTGACCTGGAAACCCACTTACCGTGTGGTCTACCGCG 1969
QY 647 AGAGTTACTTGGCGTTCTCAACTCAGCGTTCTTCCCTGGACAGCTGGACAGCGCCATCC 706
DB 1970 ACAGCTATATCCGCTTTTCCACGACGCCCTTCTCCCTGAAGAACCTGGACAACTCAGTGC 2029
QY 707 ACTGTGCAACAAACCGCGTTCAGAAAGTACTGAAGAAATGATGTGGCGCGAGCCCGCTGC 766
DB 2030 ACCTGTGCAACAACTCCATCCAGAAAGCAGCTGGAGAACTCATGCCATCGGCATCCACTGC 2089
QY 767 TGCCCCCAGACACATGTCGACCAAGCAGCAGGTTCAGAGGTACCTGACGCGCCAGGGCC 826
DB 2090 TTCCGCGCAGACACATGTGTGTAGCTAGAGGTTCAGGCGCCACCTGCGAGGAGATGGGTG 2149
QY 827 GTGGCGCCGTGTGGGCGAGCGTTCATCTACCCGCTTCCATGAAGAGGCGCATCGCCAGCCA 886
DB 2150 CCCCRAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGATGCTGTGATCCAGCAC 2209
QY 887 TGAAGTGGCCCGACGACCAAGTGGAGCGCTTCGCAAGAAAGCTTTGAGCTCTACGGGGCTG 946
DB 946
```

```
DB 2210 TTCCAGACCTCCAGGACACCCGTGCACTGTGCGAAGCCAGCTTTTGAGCTCTATGGCGCTG 2269
QY 947 ACTTGTGCTCTTGGAGGGAGCTTTCCAGCCCTGCTGATCGAGATCAATTTCCAGGCCCCACCA 1006
DB 2270 ACTTGTGTTTGGGAGGAGCTTTCCAGCCCTGCTGATTTGAGATCAACGCGAGGCCACCA 2329
QY 1007 TGCACCCCGTCCACGCGGTGTGCGGGCCAGCTGTGTGCACAGGTGTCAGGAGGACACCATCA 1066
DB 2330 TGGCACCCTCCACAGCAGTCACTGCCCGCTCTGTGTGCTGGCGTGCAGAGCTGACACCCCTGC 2389
QY 1067 AGTGGCC-----GTGAGCCGAGCTGTGACATCGCAACTTTGAGCTCTCTGT 1114
DB 2390 GGTGTGTCATTGACCGGAGGCTGGACCCCACTGTGACACAGGAGCTTTGAGCTCATCT 2449
QY 1115 GGAGCGAGCGGTGTGAGCGCGCCCATTCAGCGGGTCCGACCTCTGCTGGCGGGCGG 1174
DB 2450 ATGAGCAGCTCTGCTGTGAGGTGCTCAATATGTGGGCATCGGCTCTCTGGTAGAGGCT 2509
QY 1175 TCAGTGTGTGAGGAGAGCCAGG 1194
DB 2510 TCACCATCAAGAAGCCCATG 2529

RESULT 8
HSM800637      2553 bp      mRNA      linear      PRI 18-FEB-2000
LOCUS          Homo sapiens mRNA; cDNA DKFZp434B103 (from clone DKFZp434B103);
DEFINITION     partial cds.
ACCESSION      AL096725
VERSION        AL096725.1 GI:5419858
KEYWORDS       Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 2553)
AUTHORS        Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE          Submitted (16-JUN-1999) MIPS, Am Klopferspitz 18a, D-82152
JOURNAL        Martinried, GERMANY
COMMENT        Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
                Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
                sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
                consortium of the German Genome Project.
                This clone (DKFZp434B103) is available at the RZPD in Berlin.
                Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
                Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
                information about the clone and the sequencing project is available
                at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES             Location/Qualifiers
     1..2553
         organism=Homo sapiens
         mol_type=mRNA
         db_xref="RZPD:DKFZp434B103"
         db_xref="taxon:9606"
         clone="DKFZp434B103"
         tissue type="testis"
         clone_lib="434 (synonym: htes3). Vector pSport1; host
         DH10B; sites NotI + SalI"
         dev_stage="adult"
         1..2553
         gene="DKFZp434B103"
         892..1950
         CDS
         gene="DKFZp434B103"
         note="similarity to tubulin-tyrosine ligase"
         codon_start=1
         product="hypothetical protein"
         protein_id="CAB46375.1"
         db_xref="GI:5419859"
         db_xref="GOA:Q9Y4R7"
         db_xref="UniProt/Swiss-Prot:Q9Y4R7"
         translation="MDIDKDLAPLYLTPGKSLFLQRYVYVHGAELRLDQVOR
         CEDILQQLQAVVPOIDMEGRNIWIYKPKAKSRGRGIMCMHLEMLKLIVNGPVVMK
```


Db 1703 CCCTCCTCATCTTTGGCACCAGTTTGACCTCAGACAGTGGTTCTCGTAACTGACTGG 1762
Qy 619 AACCCCTGACCACTCGTTTCTACAAGGAGAGTTACTTGGCGTTCTCAAACTCAGCGCTTC 678
Db 1763 AACCCACTTACCGTGTGGTTCTACCGGACAGCTATATCGCTTTTCCAGCGAGCCCTTC 1822
Qy 679 TCCTTGACAGTGTGACAGCGGCATTCACCTGTGTGCAACAGCGCGTCCAGAGTACCTG 738
Db 1823 TCCTTGAAGAACCTGACAACTCAGTGCACCTGTGCAACAACTCCATCCAGAAGCACCTG 1882
Qy 739 AAGAATGATGTGGCGCGCAGCCCTGCTGCCGCACACAAACATGTGGACCGACCGAG 798
Db 1883 GAGAACTCATGCCATCGGCATCCACTGCTTCCGCCAGACAAACATGTGGTCTAGCCAGAG 1942
Qy 799 TTCCAGGAGTACCTGACGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTCATCTACCG 858
Db 1943 TTCCAGGCCACCTGCAGGAGATGGTGGCCCCAAATGCTTGGTCCACCATCATCTGTCCT 2002
Qy 859 TCCATGAAGAGCCATCGCCACAGCCATGAAGTGGCCAGGACCAACATGTGGAGCCTCGC 918
Db 2003 GGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGACACACCGTGCAGTGTCC 2062
Qy 919 AAGAACAGCTTTGAGCTCTACGGGCTGACTTGCCTTTGGGAGGAGCTTCAGGCCCTGG 978
Db 2063 AAGGCCAGCTTTGAGCTCTATGCGCTGACTTCTGTTTGGGAGGAGCTTCAGGCCCTGG 2122
Qy 979 CTGATCGAGATCAATTCACGCCCCACCATGCACCCCGTCCAGCGCGTCCAGGCGCCAGCTG 1038
Db 2123 CTGATTGAGATCAACGCCAGCCCCAGATGGCACCCCTCCACAGCAGTCACTGCCGCTC 2182
Qy 1039 TGTGACAGGTGACGAGAGACACCATCAAGGTGGCC-----GTGACCGCGAC 1086
Db 2183 TGTGCTGGGTGCAAGCTGACACCTCGCGCTGGTCAATTGACCGGATGTGGACCGCAAC 2242
Qy 1087 TGTGACATCGGAACCTTCAGCTTCTGTGAGGCGACCGCT 1127
Db 2243 TGTGACAGAGGACCTTTGAGCTCATCTATAAGCAGCCCGT 2283

RESULT 10
AX834642 2848 bp DNA linear PAT 15-DEC-2003
LOCUS Sequence 1766 from Patent EP1347046.
DEFINITION AX834642
ACCESSION AX834642
VERSION AX834642.1 GI:39920777
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Negahari, K. and Masuno, Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1347046-A 1766 24-SEP-2003;
Research Association for Biotechnology (JP)
FEATURES Location/Qualifiers
source 1..2848
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 25.8%; Score 418; DB 6; Length 2848;
Best Local Similarity 63.9%; Pred. No. 1.9e-61;
Matches 648; Conservative 0; Mismatches 365; Indels 1; Gaps 1;
Qy 80 GGGAGGAGCGCGGAGCAGCGACCTCAGCAGCAGGCAAGATGCTGAAAATGCTGAGGCAA 139
Db 1213 GGGATCGTACAGGCCCTCCCTTATGTGACGAGACACAGCCCAAGAAACAGGAGAA 1272

Qy 140 AGCTCAG--GGGCTCCCGGGGCGAGCTTGTGTGACATCGCGGTGCAAGGTGTGCGAGCGCTAC 198
Db 1273 AACCCAGTGTGGTGTCCCCAGAGTTTGTGATGAAGCTCTGTGTGCGTGCAGGAGTAC 1332
Qy 199 CTGGGGCAGCTCGGACATGAGGACATCGACAGTCCAGCAGTGCAGGATGCGTGTGAGAGCACTCACT 258
Db 1333 CTTAGCAACTTGGCCCAATGGACATCGACAGGACCTGGAGGCCCGCTGTACTCTCACC 1392
Qy 259 GAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGCGGATGCTTTTC 318
Db 1393 CCCGAGGGCTGTCTCTTCCCTCCAGCGCTACTACCAAGTGTCCACGAAGGGGCGAGAA 1452
Qy 319 ATCTCCAATTCAAGAAAATTACTTTTCGAGTGCAGGCTCTCTGCTGAATAGAAATCAAGTCT 378
Db 1453 CTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTTCGACGAGCTGCAGGCC 1512
Qy 379 GTGAACCCCTCAGACGAGCAATTGACGGGCTCCGGAACATCTGATTATTAAGCCCGGGCC 438
Db 1513 GTGGTACCCAGATAGACATGGAAGGGGATCCGAACATCTGGATCGTGAAGCCAGAGCC 1572
Qy 439 AAGTCCCGGGGCGAGACATAGTGTGCATGGAACCGTGTGGAGGAGATCTCGAGCTGGCA 498
Db 1573 AAGTCCCGTGGACGAGGCAATTATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGTG 1632
Qy 499 GGTGACAGACACCTCTTTCCAGGAGCAACAAAGTGGTGGTCCAGAAATGACATCGAGAGC 558
Db 1633 AACGGCAACCCCGTGTGATGAAGGACGCGCAAGTGGTGTGCAAGATATATTGAGCG 1692
Qy 559 CGCGTCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTGTCTACGAGCTGG 618
Db 1693 CCCCCTCTCATCTTTTGGCACCAAGTTTACCTCAGACAGTGGTTCCTGTAACTGACTGG 1752
Qy 619 AACCCCTCGACCATCTGTTTCTACAAGGAGAGTTACTTTCGCGTTTCTCAACTCAGCGCTTC 678
Db 1753 AACCCACTTACCGTGTGGTTCTACCGGACAGCTATATCCGCTTTTCCACGCGAGCCCTTC 1812
Qy 679 TCCCTGGACAGCTGGACAGCGCCATCCACTGTGTGCAACAGCGCGTCCAGAAATGACTGT 738
Db 1813 TCCCTGAAGAACCTCGACAACTCAGTGCACCTGTGCAACAACTCCATCCAGAAGCACCTG 1872
Qy 739 AAGAATGATGTGGCGCGCAGCCCCCTGTCGCCGACACAACTGTGGACAGCAGCCAGG 798
Db 1873 GAGAACTCATGCGCATCGGCATCCACTGTTCCGCCAGACAACTGTGTCTAGCCAGAG 1932
Qy 799 TTCCAGGAGTACTCTGACGCGCGCAGCGCGTGTGGCGCGTGTGGGGCAGCGTCACTTACC 858
Db 1933 TTCCAGGCCACCTCGCAGGAGATGGTGGCCCCAAATGCTTGGTCCACCATCATCTGCTC 1992
Qy 859 TCCATGAAGAGGCCATCGCCACGCCATGAAGTGGCGCCAGGACCAACGTTGAGCGCTGC 918
Db 1993 GGCATGAAGGATGTGTGTATCCACGCACCTTCAGAGCTCCCGAGCACCGTGCATGTCTCG 2052
Qy 919 AAGAACAGCTTTGAGCTCTACGGGCTGACTTTCGTCCTTTGGGAGGAGCTTTCAGGCCCTGG 978
Db 2053 AAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCTGTGTTCGGGGTGGACTTTCAGGCCCTGG 2112
Qy 979 CTGATCGAGATCAATTCCAGCCCCACCATGCAACCGCTCCACCGCGTCCAGCCCGCAGCTG 1038
Db 2113 CTGATTGAGATCAACGCCAGCCCCACGATGGCACCTCCACAGCAGTCACTGCCCGGCTC 2172
Qy 1039 TGTGACAGGTGTCAGGAGGACACCATCAAGTGTGGCGCGTGGAGCCGCGAGCTGTGTAC 1092
Db 2173 TGTGTGGCGTGCAGAGCTGACACCCCTGGCGGTGGTCAATTGACCGGAGGTGGAC 2226

RESULT 11
AK097236 2848 bp mRNA linear PRI 30-JAN-2004
LOCUS Homo sapiens cDNA FLJ39917 fis, clone SPLEN2019405, highly similar
DEFINITION to Homo sapiens FHTTL protein mRNA.
ACCESSION AK097236
VERSION AK097236.1 GI:21756925
KEYWORDS oligo capping; fis (full insert sequence).

SOURCE	Homo sapiens (human)	ORIGIN	Query Match
ORGANISM	Homo sapiens		Best Local Similarity 25.8%; Score 418; DB 9; Length 2848;
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		Matches 648; Conservative 0; Mismatches 365; Indels 1; Gaps 1;
AUTHORS	Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayaishi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yanazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hoto, T., Kusano, J., Kanehori, K., Takahashi, F., A., Hara, H., Tanase, T., Nomura, Y., Toqiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hiki, J., Kobatake, N., Inagaki, H., Ikema, Y., Shigeta, K., Senba, T., Matsumura, K., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.		
TITLE	Complete sequencing and characterization of 21,243 full-length human cDNAs		
JOURNAL	Nat. Genet. 36 (1), 40-45 (2004)		
PUBMED	14702039		
AUTHORS	2		
REFERENCE	Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 2848)		
AUTHORS	Isogai, T. and Yamamoto, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (B-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.		
FEATURES	Location/Qualifiers		
source	1. 2848 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="SPLEN2019405" /tissue_type="spleen" /clone_lib="SPLEN2" /note="Cloning vector: pME18SFL3"		

```

RESULT 12
AF078842
LOCUS      1684 bp      mRNA      linear      PRI 09-JAN-2000
DEFINITION Homo sapiens HOTTTL protein mRNA, complete cds.
ACCESSION AF078842
VERSION    AF078842.1 GI:6683744
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1684)
AUTHORS   Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
            Zhou, S., Liu, M. and He, F.
TITLE     Functional prediction of the coding sequences of 50 new genes
            deduced by analysis of cDNA clones from human fetal liver
            Unpublished
JOURNAL   2 (bases 1 to 1684)
REFERENCE Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
            Zhou, S., Liu, M. and He, F.
AUTHORS   Direct Submission
TITLE     Submitted (17-JUL-1998) Dept. of Experimental Hematology, Beijing
            Institute of Radiation Medicine, 27 Taiping Rd, Beijing 100850,
            P.R.China
FEATURES
            Location/Qualifiers
            1..1684
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="HQ0207"
               /tissue_type="liver"
               /dev_stage="fetus"
            1..398
               /codon_start=1
            399..1277
               /notes="hnp0207; similar to bovine and pig tubulin-tyrosine
               ligase (TTL): Swiss-Prot Accession Numbers P38584 and
               P38160"
            5'UTR
            CDS
            3'UTR
            ORIGIN
            Query Match      23.4%; Score 380.2; DB 9; Length 1684;
            Best Local Similarity 68.6%; Pred. No. 5.6e-55;
            Matches 545; Conservative 0; Mismatches 238; Indels 12; Gaps 1;

345 GCAGTCCAGGCTCTGCTGAATAGATCAGCTGTGTGAACCTTCACAGCGACATTGACGG 404
347 GCGCTGTGAGGACATCTTCAGCAGCTGCGAGCGGTGTATCCCAATAGATGAGGAAG 406
405 GCTCCGGAACATCTGATTATTAAGCCCGCGGCCCAAGTCCCGGGCCGAGACATAGTGTG 464
407 GGATCGCAACTTCTGATCTGTGAAGCCAGGAGCAAGTCCCGCGGAGGAGGCATCATGTG 466
465 CATGGACCGTGTGGAGGAGATCTCGAGCTGGCAGCTGCAGACCACCTCTTTCAGGGA 524
467 CATGGACCACTCTGGAGGAGATGCTGAAGCTGGTGAACCGCAACCCCGTGTGATGAAGA 526
525 CAACAAGTGGGTGGTCCAGAAGTACATCGAGACCGCGCTGCTCATCTGTGACACCAAGTT 584
527 CGGCAAGTGGGTGGTCCAGAAGTATATTGAGCGGGCCCTCTCTCATCTTTGGCACCAAGTT 586

```

TITLE

Generation and initial analysis of more than 15,000 full-length

```

QY      585 CGACATCAGACAGTGTTCCTCGTCA CGGACTGGAACCCCTTGACCATCTGGTTCTACAA 644
DB      587 TGACCTCAGACAGTGTTCCTCGTAACTGACTGGAACCCACTTACGTTGGTTCTACCG 646
QY      645 GGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGGACAGCTCGACAGCGCCAT 704
DB      647 CGACAGCTATATCCGCTTTTCCACGAGCCCTTCTCCCTGAAAGAACCTTGGAACTCAGT 706
QY      705 CCACCTGTGCAACAAACCGCGTCCAGAACTACCTCGAAGATGATGTGGCGCGCAGCCCTCT 764
DB      707 GCACCTGTGCAACAACTCCATCCAGAGCACTCGGAGAACTCATGCCATCGGCATCCACT 766
QY      765 GTGCCCGGACACAAATGTGGAACAGCAGCAGGTTCCAGGAGTACCTGACAGCGCAGGG 824
DB      767 GCTTCGGCCAGACAAATGTGCTAGCCAGAGGTTCCAGGCGCCACCTGACAGAGATGG 826
QY      825 CGGTGGCGCGTGTGGGCGAGCGTCTATCTACCGTCCATGACCGTCCATGAAGAGGCCATCGCCACGC 884
DB      827 TGCCCAAAATGCTTGGTCCACCATCATGTGCTGGCATGGAAGGATGCTGTGATCCACGC 886
QY      885 CATGAAGTGGCGCCAGGACCACTGGAGCCTCGCAAGAACAGCTTTAGCTCTACGGGCG 944
DB      887 ACTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGCCAGCTTTAGCTCTATGGCG 946
QY      945 TGACTTGCTGTGGAGGAGACTTCAGGCCCTGGTGTGATGAGATCAATTCAGGCGCCAC 1004
DB      947 TGACTTGTGTGGGAGGAGACTTCAGGCCCTGGTGTGATGAGATCAACGCCAGCGCCAC 1006
QY      1005 CATGACCGCTCCAGCGCGTCCAGCGCCAGCTGTGTCACAGGTGTCAGGAGGAGACCAT 1064
DB      1007 GATGGACCCCTCCAGCAGTCACTGCGCGGCTCTGTGTGGCGTGCAGCTGACACCT 1066
QY      1065 CAAGGTGGCC-----GTGGACCGCAGCTGTGACATCGGCAACTTCGAGTCTCT 1112
DB      1067 GCGCGTGGTCATTGACCGGATGCTGACCGCAACTGTGACAGGAGCCTTTGAGTCTAT 1126
QY      1113 GTGAGGCGAGCGCGT 1127
DB      1127 CTATAAGCAGCGCGT 1141

RESULT 13
BC006830
LOCUS
DEFINITION IMAGE:3597662), complete cds.
ACCESSION BC006830
VERSION    BC006830.1 GI:13905089
KEYWORDS   MGC.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1897)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,P.,
            Diatchenko,L., Marusina,K., Farmer,A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fanej,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalek,U., Smalley,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
            Generation and initial analysis of more than 15,000 full-length

```

human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
2388257
12477932
2 (bases 1 to 1897)
Direct Submission
Strausberg, R.
Submitted (27-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: scapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 16 Row: g Column: 10.
Location/Qualifiers
1. .1897
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:11856 IMAGE:3597662"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI CGAP Mam6"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. .1897
/gene="4833441J24Rik"
/db_xref="LocusID:101100"
/db_xref="MGI:2141418"
590. .1390
/gene="4833441J24Rik"
/codon_start=1
/product="4833441J24Rik protein"
/protein_id="AAH06830.1"
/db_xref="GI:13905090"
/db_xref="LocusID:101100"
/db_xref="MGI:2141418"
/translation="MEGDNIWIKPKAKSRGIMCMNPLDEMLKLVDCNPLMKDGS
KWIQYIERPLIFGKTLRQWFLVDNPLTVFYRDSYRFTQPSLKNLDS
VHLNNSIQHLASCHRRPMLPDPNWSRFOAHLQVEDDAPKWSVIVPMKAIV
IHALQTSQDNVQRKAFELYGADFVGEDFQPLIEINASPTMAFSTAVRLCAGV
QADTLRVVIDRDRSCDTGAFELIYQVRCSSTSHKQPHGPGIWCTE"
590. .1270
/gene="4833441J24Rik"
/note="TTL; Region: Tubulin-tyrosine ligase family.
Tubulins and microtubules are subjected to several
post-translational modifications of which the reversible
deetyrosination/tyrosination of the carboxy-terminal end of
most alpha-tubulins has been extensively analysed. This
modification cycle involves a specific carboxypeptidase
and the activity of the tubulin-tyrosine ligase (TTL). The
true physiological function of TTL has so far not been
established. Tubulin-tyrosine ligase (TTL) catalyses the
ATP-dependent post-translational addition of a tyrosine to
the carboxy terminal end of deetyrosinated alpha-tubulin.
In normally cycling cells, the tyrosinated form of tubulin
predominates. However, in breast cancer cells, the

deetyrosinated form frequently predominates, with a
correlation to tumour aggressiveness. On the other hand,
3-nitrotyrosine has been shown to be incorporated, by TTL,
into the carboxy terminal end of deetyrosinated
alpha-tubulin. This reaction is not reversible by the
carboxypeptidase enzyme. Cells cultured in 3-nitrotyrosine
rich medium showed evidence of altered microtubule
structure and function, including altered cell morphology,
epithelial barrier dysfunction, and apoptosis"
/db_xref="CDD:pfam03133"

ORIGIN

Query Match	22.6%	Score 366.4	DB 10	Length 1897
Best Local Similarity	59.3%	Pred. No. 1.2e-52		
Matches 499	Conservative 0	Mismatches 221	Indels 0	Gaps 0

QY	379	GTGAACCTCTCAGACGACATTTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCC	438
DB	572	GTGGTACCCAGTTAGACATGGAGGGGATCAGAACATCTGGATCGTGAAGCCCGGAGCC	631
QY	439	AAGTCCCGGGCCGACACATAGTGTGATGGACCGTGTGGAGGAGATCTCTGGAGCTGGCA	498
DB	632	AAGTCCCGGGCCGACGAGATTTATGTGCATGAACCGCTCGATGAGATGCTTGAAGCTGGT	691
QY	499	GCTGCAGACCACTCTTCCAGGGACAACAGTGGTGTCCAGAGTACATCGAGACG	558
DB	692	GACTGCAACCCCATGCTCATGAAGATGGCAAGTGGATCGTGAGAGTACATTTAGCGG	751
QY	559	CCGCTGCTATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGCGACTGG	618
DB	752	CCCTGCTTATCTTCGGCACCAAGTTTGACCTGAGACAGTGGTTCCTGGTGAAGTGGT	811
QY	619	AACCCCTGACCACTTGGTTTACAAGAGAGTTACTTGGGGTTCCTCAACTCAGCGCTTC	678
DB	812	AACCCACTACCTGTGGTTTACCGAGACAGCTACATTCGCTTCTCCACACAGCCCTTC	871
QY	679	TCCTGGACAAGCTGACAGCGCCATCCACCTGTGTGCAACAACCGCTGCCAGAGTACTCG	738
DB	872	TCCTGGAAGAACCTTGACAACTCTGTGCACCTGTGTAAACAACCTCCATCCAGACACTT	931
QY	739	AAGAATGATGTGGCCGCGAGCCCTCTGCTGCCCGCACACAACATGTGGACAGCAGCCAGG	798
DB	932	GAGGCTTCTGTGTCACCGCACCCGATGCTGCCCCAGATAACATGTGGTCCAGCCAGAGG	991
QY	799	TTCCAGAGTACTGTGAGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTCATCTACCCG	858
DB	992	TTTCAGGCCACTTTCAGGAGGTAGATGCCCCAAAGGCCCTGTGTCAGCGCTCATCTGTC	1051
QY	859	TCCNTGAAGAGCCATCGCCACGCCATGAGGTGGCCGAGGACCAACCGTGGAGCCCTCG	918
DB	1052	GGCATGAAGGCTGCTGTGATCCACGCCCTGACACCTCCCAAGAACACGTCGAGTGC	1111
QY	919	AAGAACAGCTTTGAGCTCTACGGGGTGAATTCCTTGGGAGGAGCTTCAGGCGCTCG	978
DB	1112	AAGCCAGCTTTGAGCTCTATGGGGCAGACTTTGTGTTGGGGAAGACTTCCAGCCCTG	1171
QY	979	CTGATCGAGATCAATTCAGCCCCACCATCCACCCCGTCCACCGCGGTCAAGCGCCAGCTG	1038
DB	1172	TTGATTGAAATCAATGCCAGCCCCACCATGGCACCTTCACAGGCTGTCACTGCCCCCTC	1231
QY	1039	TGTGCAGAGTGCAGAGGACACCATCAAGGTGGCGCTGGACCGCAGCTGTGACATCGG	1098
DB	1232	TGTGCGCGTGTGCAGGACAGATACCCCTGCGTGTGTGTCATCGACCGCGCTGGGACCG	1291

RESULT 14

AC119959	AC119959	163783 bp	DNA	linear	ROD 23-JUL-2004
LOCUS	Mus musculus chromosome 15, clone RP24-467H19, complete sequence.				
DEFINITION					
ACCESSION	AC119959				
VERSION	AC119959.8	GI:50540805			
KEYWORDS	HTG.				
SOURCE	Mus musculus (house mouse)				

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 163783)

AUTHORS

Birren B., Nusbaum, C. and Lander, E.

TITLE

Mus musculus chromosome 15, clone RP24-467H19

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 163783)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lanazares, R., Landers, T., Lebecky, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Ribback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (02-MAY-2002)

Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 163783)

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (26-MAY-2004)

Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

4 (bases 1 to 163783)

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (23-JUL-2004)

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Jul 23, 2004 this sequence version replaced gi:47679219.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L25717

Center clone name: 467_H_19

FEATURES

Location/Qualifiers

1..163783

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP24-467H19"

/clone_lib="RPCI-24 Male Mouse BAC"

1..4

/note="clone boundary"

clone end:SP6

site:Mbol

complement(169..204)

/rpt_family="ORR1B1"

818..994

/rpt_family="ORR1D"

1312..1527

/rpt_family="B3"

1573..1645

/rpt_family="ID3"

1665..1793

/rpt_family="B1P"

1794..1959

/rpt_family="B3"

complement(2256..2658)

/rpt_family="LI"

complement(2659..2833)

/rpt_family="B2_Mm2"

complement(2834..3302)

/rpt_family="LI"

complement(3311..3448)

/rpt_family="B1_MM"

complement(3449..4218)

/rpt_family="LI"

4214..4557

/rpt_family="Lx6"

5120..5279

/rpt_family="CT-rich"

complement(5297..5423)

/rpt_family="B1P"

complement(5451..5524)

6016..6494

/rpt_family="RLTR11A"

complement(6678..6797)

/rpt_family="B1_MM"

complement(6810..6945)

/rpt_family="B1_MM"

7008..7148

repeat_region

repeat_region	/rpt_family="B1_MM"
complement(7763. .7941)	
repeat_region	/rpt_family="B3"
complement(7960. .8105)	
repeat_region	/rpt_family="B1_MM"
complement(8121. .8300)	
repeat_region	/rpt_family="B3A"
8486. .8625	
repeat_region	/rpt_family="B4A"
8618. .8788	
repeat_region	/rpt_family="B1_MM"
8804. .9102	
repeat_region	/rpt_family="L1MB2"
9162. .9217	
repeat_region	/rpt_family="PB1D7"
complement(9339. .9460)	
repeat_region	/rpt_family="B3"
complement(9598. .9748)	
repeat_region	/rpt_family="B3"
complement(9755. .12502)	
repeat_region	/rpt_family="L1_MM"
12503. .12992	
repeat_region	/rpt_family="IAPLTR2_MM"
12994. .13717	
repeat_region	/rpt_family="IAPLTR2_MM-int"
13718. .13836	
repeat_region	/rpt_family="IAPLTR2_MM-int"
13832. .14083	
repeat_region	/rpt_family="IAPLTR2_MM-int"
14083. .15904	
repeat_region	/rpt_family="IAPLTR2_MM-int"
15905. .16394	
repeat_region	/rpt_family="IAPLTR2_MM"
complement(16395. .20012)	
repeat_region	/rpt_family="L1"
complement(20013. .20157)	
repeat_region	/rpt_family="B1_MM"
20157. .20290	
repeat_region	/rpt_family="B2_Mm2"
20410. .20553	
repeat_region	/rpt_family="B1_MM"
20586. .20777	
repeat_region	/rpt_family="B2_Mm2"
21012. .21082	
repeat_region	/rpt_family=" (CCAA)n"
21132. .21294	
repeat_region	/rpt_family="PB1D10"
21301. .21342	

Query Match	17.5%;	Score 283.6;	DB 10;	Length 163783;
Best Local Similarity	76.7%;	Pred. No. 7.3e-39;		
Matches 365;	Conservative 0;	Mismatches 99;	Indels 12;	Gaps 1;
Qy	585	GACAACTGGACAGCGCCATCCACCTGTGTCCAAACACGCCGTCAGAGTACTCTGAAGAT	744	
Db	136973	GACATTGCCCCACAGCGCCATCCACCTGTGTAACTCCATCCAGAGCGCCTCAAAAT	137032	
Qy	745	GATGTGGSCGACGCCCCCTGCTGCTCCGCACACAACTGTGTGACACGACACCAAGTTCCAG	804	
Db	137033	GATAAAGAGCGCAGTCCCGCTGCTACTTTGTCAATCATGTGTGACCACTCGTTTCCAG	137092	
Qy	805	GAGTACCTGCAGCGCCAGCGCCGTGTCGGCGCAGCGGTCTATCTACCCGCTCCATG	864	
Db	137093	GAGTACCTGCAGAGAGGGGCGAGGAGCACTGGGGTAGGCATCATCTACCCGCTCATG	137152	
Qy	865	AAGAAGGCCATCGCCCCACGCCATGAAGTGGGCCACAGGACCACTGTGGAGCCTCGCAAGAC	924	
Db	137153	AAAAGAGCTGTCAACCAATGCCATGAGGGTAGCCACAGGACCACGTAGAAGCCGCTAAGAC	137212	
Qy	925	AGCTTTGAGCTCTACGGGGCTGACTTCGTCTCTTTGGAGGGGACTTCAGGCCCTGGCGTGATC	984	
Db	137213	AGCTTCGAGCTCTACGGAGCCGACTTCATCTCTGGGGCGAGACTTCAAGCCATGGCTTATC	137272	

QY	985	GAGATCAATTCCAGCCCCACCATGACCCCGTCCACGCCCGGTACGCCGCCAGCTGTGTGCA	1044
Db	137273	GAGATCAATCTCAGCCCCACCATGACCCCGTCCACCTCCCGTCCACAGCTCAGCTGTGTGCC	137332
QY	1045	CAGGTGCAGGAGGACACCATCAAGGTGGCCG-----TGCACCGCAGCTGTGTGAC	1092
Db	137333	CAGGTGCAGGAGGACACCATCAAGGTGTGTGGATCGCAAACTGCACCAACTGTGTGAC	137392
QY	1093	ATCGGCAACTTCGAGCTCCTGTGGAGGACGCCGGTGGTGTGAGCCGCCCCCATTCAG	1148
Db	137393	ATCGGCAACTTTGAGTCTCTGTGGCGCAGGTGAGGATAGCAGCTACCCCTTCTTAG	137448
RESULT	15		
AC117700		201420 bp	DNA
LOCUS		Mus musculus chromosome 15	clone RP23-455J20 map 15, *** SEQUENCING
DEFINITION		IN PROGRESS ***	3 unordered pieces.
ACCESSION	AC117700		
VERSION	AC117700.17	GI:52694698	
KEYWORDS	HTG; HTGS PHASE1; HTGS FULLTOP; HTGS ACTIVEFIN.		
SOURCE	Mus musculus	(house mouse)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1	(bases 1 to 201420)	
AUTHORS	Birren, B., Nusbaum, C. and Lander, E.		
TITLE	Mus musculus chromosome 15, clone RP23-455J20		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 201420)	
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,		

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boquelavkiy, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Chospel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gort, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McSwan, P., McKernan, K., Meldrim, J., Meneus, L.,
Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201420)

Barren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavkiy, L., Boukhgalter, B., Camarata, J., Chang, J.,
Chospel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodor, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201420)

Barren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavkiy, L., Boukhgalter, B., Camarata, J., Chang, J.,
Chospel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodor, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201420)

Barren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavkiy, L., Boukhgalter, B., Camarata, J., Chang, J.,
Chospel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodor, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201420)

Barren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavkiy, L., Boukhgalter, B., Camarata, J., Chang, J.,
Chospel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodor, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201420)

Barren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,
Bloom, T., Boguslavkiy, L., Boukhgalter, B., Camarata, J., Chang, J.,
Chospel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodor, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201420)

Barren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barn

Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-SEP-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 26, 2004 this sequence version replaced gi:52077711.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L24098

Center clone name: 455_J_20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 82149: contig of 82149 bp in length
* 82150 82249: gap of unknown length
* 82250 98003: contig of 15754 bp in length
* 98004 98103: gap of unknown length
* 98104 201420: contig of 103317 bp in length.

FEATURES

Location/Qualifiers
1..201420
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-455J20"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match 17.5%; Score 283.6; DB 2; Length 201420;
Best Local Similarity 76.7%; Pred. No. 7.1e-39;
Matches 365; Conservative 0; Mismatches 99; Indels 12; Gaps 1;

QY 685 GACAAAGCTGACAGCGCCATCCACCTGTGTCAACAACGCCGTCCAGAAAGTACCTGAAGAAT 744
Db 4228 GACATTGCCACAGCGCCATCCACCTGTGTCAACAACGCCGTCCAGAAAGTACCTGAAGAAT 4287

QY 745 GATGTGGCGCGAGCCCGCTGCTGCCGACACAACATGTGGACAGCAGCCAGGTTCCAG 804
Db 4288 GATAAAGAGCGCAGTCCGCTGTCTACCTTGTATACATGTGGACAGCAGCAGTCCGTTCCAG 4347

QY 805 GAGTACCTGACGCGCCAGCGCGCTGTGTGGGCGAGCGCTCATCTACCGCTCCATG 864
Db 4348 GAGTACCTGACGAGAGGGGCGCAGGAGGACGTTGGGTAGCATCATCTACCGCTCATG 4407

QY 865 AAGAAGGCCATCGCCCAACGCCATGAAGGTGGCCAGGACCAACGTTGAGCGCTCGCAAGAAC 924
Db 4408 AAAAGAGCTGTCAACCAATGCCATGAGGGTAGCCAGGACCCAGGAGCCCGTAGAAGCCCGTAAGAAC 4467

QY 925 AGCTTTGAGCTCTACGGGGTGACTTCTGCTTGGGAGGACTTCAGGGCCTGGCTGATC 984
Db 4468 AGCTTGAAGCTCTACGGAGCGGACTTCTATCCTGGGGCGAGACTTCAAGCCATGGCTTATC 4527

QY 985 GAGATCAATTCAGCGCCCAACCATGACCCCGCTCCACGCCGGTACCGGCCAGCTGTGTGCA 1044
Db 4528 GAGATCAATTCAGCGCCCAACCATGACCCCGCTCCACGCCGGTACCGGCCAGCTGTGTGCC 4587

QY 1045 CAGGTGAGGAGGACACCATCAAGGTGGCGG-----TGGACCGAGCTGTGAC 1092

Db 4588 CAGGTGACGAGGACACCATCAAGGTGTGTGGATCGCAAACTGGACCGAAACTGTGAC 4647
QY 1093 ATCGGCAACTTCGAGCTCCTGTGGAGGACGCCGCTGTGAGCGCCGCCCATTTAG 1148
Db 4648 ATCGGCAACTTTGAGCTTCTGTGGCGGACGTTAGGATAGCAGCTACCCCTTCTTAG 4703

Search completed: September 25, 2005, 11:26:54
Job time : 7536.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 23, 2005, 23:48:14 ; Search time 1417.93 Seconds
(without alignments)
8666.982 Million cell updates/sec

Title: US-10-615-659-1
Perfect score: 1838
Sequence: 1 tggagcagcctggggccc.....actctcccaaggcggaattc 1838

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq: *
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq: *
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq: *
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq: *
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq: *
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq: *
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq: *
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq: *
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq: *
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1838	100.0	1838	19	US-10-615-659-1
2	1838	100.0	1838	19	US-10-615-659-1
3	1782.8	97.0	3554	19	US-10-615-659-12
4	1782.8	97.0	3554	19	US-10-615-659-12
5	1678.8	91.3	3465	19	US-10-615-659-11
6	1678.8	91.3	3465	19	US-10-615-659-11
7	1677.2	91.3	1939	19	US-10-615-659-9
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 12, Appl
					Sequence 12, Appl
					Sequence 11, Appl
					Sequence 11, Appl
					Sequence 9, Appli

8	1677.2	91.3	1939	19	US-10-615-659-9	Sequence 9, Appli
9	1574.8	85.7	1859	19	US-10-615-659-10	Sequence 10, Appl
10	1574.8	85.7	1859	19	US-10-615-659-10	Sequence 10, Appl
11	726	39.5	726	19	US-10-615-659-3	Sequence 3, Appli
12	726	39.5	726	19	US-10-615-659-3	Sequence 3, Appli
13	501	27.3	101270	20	US-10-723-860-631	Sequence 631, Appl
14	432.8	23.5	2380	18	US-10-275-595A-39	Sequence 39, Appl
15	427	23.2	450	16	US-10-029-386-24894	Sequence 24894, A
16	421.4	22.9	2538	15	US-10-102-524-1765	Sequence 1765, Ap
17	421.4	22.9	2553	21	US-10-956-157-2016	Sequence 2016, Ap
18	421.4	22.9	2553	22	US-10-756-149-1971	Sequence 1971, Ap
19	421.4	22.9	3001	15	US-10-037-270-494	Sequence 494, App
20	421.4	22.9	3001	17	US-10-117-722-494	Sequence 1765, Ap
21	418	22.7	2848	17	US-10-108-260A-1766	Sequence 1765, Ap
22	413	22.5	2611	9	US-09-925-298-199	Sequence 199, App
23	413	22.5	2611	14	US-10-102-806-199	Sequence 199, App
24	387.6	21.1	526	13	US-10-027-632-135927	Sequence 135927,
25	387.6	21.1	526	13	US-10-027-632-135928	Sequence 135928,
26	387.6	21.1	526	17	US-10-027-632-135927	Sequence 135927,
27	387.6	21.1	526	17	US-10-027-632-135928	Sequence 135928,
28	381.8	20.8	1400	21	US-10-956-157-9356	Sequence 9356, Ap
29	381.8	20.8	1673	21	US-10-956-157-4121	Sequence 4121, Ap
30	319.8	17.4	1400	21	US-10-956-157-7251	Sequence 7251, Ap
31	272.4	14.8	2241	19	US-10-615-659-27	Sequence 27, Appl
32	272.4	14.8	2241	19	US-10-615-659-27	Sequence 27, Appl
33	196	10.7	755	13	US-10-027-632-135929	Sequence 135929,
34	196	10.7	755	13	US-10-027-632-135930	Sequence 135930,
35	196	10.7	755	17	US-10-027-632-135929	Sequence 135929,
36	196	10.7	755	17	US-10-027-632-135930	Sequence 135930,
37	190.6	10.4	1958	21	US-10-956-157-4903	Sequence 4903, Ap
38	166	9.0	1728	22	US-10-450-763-5962	Sequence 5962, Ap
39	164.4	8.9	2250	22	US-10-450-763-5962	Sequence 5962, Ap
40	163.4	8.9	418	9	US-09-983-965-2032	Sequence 2032, Ap
41	131.8	7.2	492	10	US-09-918-995-35253	Sequence 35253, A
42	130	7.1	525	16	US-10-029-386-11174	Sequence 11174, A
43	122.6	6.7	436	17	US-10-242-535A-7660	Sequence 7660, Ap
44	122.6	6.7	436	18	US-10-085-783A-7660	Sequence 7660, Ap
45	113.4	6.2	279	18	US-10-424-599-119452	Sequence 119452,

ALIGNMENTS

RESULT 1
US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Query Match 100.0%; Score 1838; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGAGCAGCCTGGGGCCCATCGCTATGAGCGCGGAAGTCGGCGGTTCGGGAGCC 60
|||||

Query Match 100.0%; Score 1838; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGGGAGCAGCGCTGGGCCCCATCGGCTATGAGGGCGGGAAGTGGGGGAGCC 60
1 TGGGAGCAGCGCTGGGCCCCATCGGCTATGAGGGCGGGAAGTGGGGGAGCC 60
61 TCCGTGGCCCTGGGCCCCATCTCCAGTCCCGAGTCCCTGGCTCGACAGATAGGGCGAGGCT 120
61 TCCGTGGCCCTGGGCCCCATCTCCAGTCCCGAGTCCCTGGCTCGACAGATAGGGCGAGGCT 120
121 GTGCTGTCTTTTCAAGAGACTTCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG 180
121 GTGCTGTCTTTTCAAGAGACTTCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG 180
181 TCAGCCACCAAGAGCTGCAGCAGGAGCAGCAGAGCAAGCAAGCCAGGACAGAGGAGG 240
181 TCAGCCACCAAGAGCTGCAGCAGGAGCAGCAGAGCAAGCAAGCCAGGACAGAGGAGG 240
241 CCGGAGCAGCGACTCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAAAAGCTCAGGG 300
241 CCGGAGCAGCGACTCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAAAAGCTCAGGG 300
301 GCCTCCCGGGGAGCTTGTGGACATCGCGTGAAGTGTGCCAGGCTA CTTGGGGCAGC 360
301 GCCTCCCGGGGAGCTTGTGGACATCGCGTGAAGTGTGCCAGGCTA CTTGGGGCAGC 360
361 TGGAGCATAGGACATCGACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGGCGAGT 420
361 TGGAGCATAGGACATCGACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGGCGAGT 420
421 GGGAGGACCTGACCCAGCAGTACTCTCCCTGTTCAATGGCGATGCTTTCATCTCCAATT 480
421 GGGAGGACCTGACCCAGCAGTACTCTCCCTGTTCAATGGCGATGCTTTCATCTCCAATT 480
481 CAAGAAATTAATTTTCAGTGCCAGGCTCTGCTGAATAGAAATCACTGTGTGAACCCCTC 540
481 CAAGAAATTAATTTTCAGTGCCAGGCTCTGCTGAATAGAAATCACTGTGTGAACCCCTC 540
541 AGACGGACATGAGCGGCTCGGAAACATCTGGATATTAAGCCCGCGGCAAGTCCCGG 600
541 AGACGGACATGAGCGGCTCGGAAACATCTGGATATTAAGCCCGCGGCAAGTCCCGG 600
601 GCCGAGACATAGTGTGATGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACC 660
601 GCCGAGACATAGTGTGATGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACC 660
661 ACCCTCTTTCCAGGACCAACAAGTGGGTGGTCCAGAAAGTACATCGAGACCGCGTGTCTCA 720
661 ACCCTCTTTCCAGGACCAACAAGTGGGTGGTCCAGAAAGTACATCGAGACCGCGTGTCTCA 720
721 TCTGTGACACCAAGTTGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTCGA 780
721 TCTGTGACACCAAGTTGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTCGA 780
781 CCATCTGTTCTCAAGGAGATTTACTTGGGTTCTCAACTCAGCGGCTCTCCCTGGACA 840
781 CCATCTGTTCTCAAGGAGATTTACTTGGGTTCTCAACTCAGCGGCTCTCCCTGGACA 840
841 AGCTGGACAGCGCCATCCACTGTCGCAACACGCGTCCAGAGTACCTGAAGATATG 900
841 AGCTGGACAGCGCCATCCACTGTCGCAACACGCGTCCAGAGTACCTGAAGATATG 900
901 TGGCCCGAGCGCCCTGTGTCGCCGACACAAATGTGGAACAGCACAGGTTCCAGAGT 960
901 TGGCCCGAGCGCCCTGTGTCGCCGACACAAATGTGGAACAGCACAGGTTCCAGAGT 960
961 ACCTGACGCGCGGCGGCGGCTGTGGGCGAGGCTCATCTACCGTCCATGAGA 1020

961 ACCTGACGCGCAGGCGCGTGGCGCGTGGGGCAGCGTCATCTACCGTCCATGAAGA 1020
1021 AGGCCATCGCCCAAGCATGAAAGTGGCCAGAGACCAAGTGGAGCCTCGCAAGAACAGCT 1080
1021 AGGCCATCGCCCAAGCATGAAAGTGGCCAGAGACCAAGTGGAGCCTCGCAAGAACAGCT 1080
1081 TTGAGCTCTACGCGGCTGACTTCGCTCTTGGAGGAGACTTTCAGGCCCTGCTGATCGAGA 1140
1081 TTGAGCTCTACGCGGCTGACTTCGCTCTTGGAGGAGACTTTCAGGCCCTGCTGATCGAGA 1140
1141 TCAATTCAGCCCCACCATGACCCGTCACGCGCGGTACGCGCCACAGCTGTGTGCACAGG 1200
1141 TCAATTCAGCCCCACCATGACCCGTCACGCGCGGTACGCGCCACAGCTGTGTGCACAGG 1200
1201 TGCAGAGGACACCATCAAGGTGGCGTGGACCGCAGCTGTGACATCGGCAACTTCAGC 1260
1201 TGCAGAGGACACCATCAAGGTGGCGTGGACCGCAGCTGTGACATCGGCAACTTCAGC 1260
1261 TCTGTGGAGGAGCGGCTGTTGAGCGCCCGCATTCAGCGGCTCGACCTCTGCGCTGG 1320
1261 TCTGTGGAGGAGCGGCTGTTGAGCGCCCGCATTCAGCGGCTCGACCTCTGCGCTGG 1320
1321 CGGCGCTCAGTGTGAGGAGACCGCAGGAGCAGGTGTGCCCTGTGCAACCTCAAGGCGCT 1380
1321 CGGCGCTCAGTGTGAGGAGACCGCAGGAGCAGGTGTGCCCTGTGCAACCTCAAGGCGCT 1380
1381 CGGCGCTCGTGTGGAGCGCGCAGCGCTGAAGGCAAGCGGCGCCCTCGGCCATGCGCAGAC 1440
1381 CGGCGCTCGTGTGGAGCGCGCAGCGCTGAAGGCAAGCGGCGCCCTCGGCCATGCGCAGAC 1440
1441 CTCGCCAGGAGACCCCATCAACAGCTCTCAGCGGACTTGGGACTGAAGGAAGAGAGG 1500
1441 CTCGCCAGGAGACCCCATCAACAGCTCTCAGCGGACTTGGGACTGAAGGAAGAGAGG 1500
1501 GGCTCCCGCTGGCGCTTGTGGCAACCTTAAGGGGGGCGAGCGGCTGGAGCGCGCAC 1560
1501 GGCTCCCGCTGGCGCTTGTGGCAACCTTAAGGGGGGCGAGCGGCTGGAGCGCGCAC 1560
1561 AGCCCAACCGCGACCAAGAGCTGTGGAGAGTGTGGAGCTCCCGGCTGCGCTGTGCGCACG 1620
1561 AGCCCAACCGCGACCAAGAGCTGTGGAGAGTGTGGAGCTCCCGGCTGCGCTGTGCGCACG 1620
1621 TGGACAGTCAAGCGCCCAACACCGGCTGCCCTAGCCAGCGCCGCAAGCTGGGATC 1680
1621 TGGACAGTCAAGCGCCCAACACCGGCTGCCCTAGCCAGCGCCGCAAGCTGGGATC 1680
1681 CAAACAGCTTAATGCGCACCCGCTGTGCTGGAGGCTGTGCTGGGGGCTGAAGACAGCAGG 1740
1681 CAAACAGCTTAATGCGCACCCGCTGTGCTGGAGGCTGTGCTGGGGGCTGAAGACAGCAGG 1740
1741 GCGCGTGTGCTCGCGCCCGGAGGAAAGGTTTCATGACAGCGTCAGATTTCTCTGACGA 1800
1741 GCGCGTGTGCTCGCGCCCGGAGGAAAGGTTTCATGACAGCGTCAGATTTCTCTGACGA 1800
1801 GGAGTACAGGTTTCAGCGCACTCTCCCAAGGGCGAATTC 1838
1801 GGAGTACAGGTTTCAGCGCACTCTCTCCCAAGGGCGAATTC 1838

RESULT 3
US-10-615-659-12
; Sequence 12, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; CURRENT APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: PatentIn version 3.2									
; SEQ ID NO 12									
; LENGTH: 3554									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-10-615-659-12									
Query Match 97.0%; Score 1782.8; DB 19; Length 3554;									
Best Local Similarity 98.3%; Pred. No. 0;									
Matches 1824; Conservative 0; Mismatches 2; Indels 30; Gaps 1;									
QY	1	TGGAGCAGCCCTGGGCCCATCGGTATGAGGGCGGGAAGTGGGCGGGTTGGGAGCC	60						
DB	1513	TGGGAGCAGCCCTGGGCCCATCGGTATGAGGGCGGGAAGTGGGCGGGTTGGGAGCC	1572						
QY	61	TCCGTGGCCCTGGCCCATCTCCAGTCCCAGTCCCTGGCTCGGACAGATAGGCGGAGGCT	120						
DB	1573	TCCGTGGCCCTGGCCCATCTCCAGTCCCAGTCCCTGGCTCGGACAGATAGGCGGAGGCT	1632						
QY	121	GTGCTGTCTCTTCA-----GAAGACTTTCGGCGCA	150						
DB	1633	GTGCTGTCTCTTTCAGAGTAGCAGCGTGGGAAGAGGATTCTCTGGAAGACTTCCGGCGCA	1692						
QY	151	CGATGGCATCCAGCATCTCAAGTGGGTGTAGGCCACGACGAGCTCGAGGAGGAGCA	210						
DB	1693	CCATGGCATCCAGCATCTCAAGTGGGTGTAGGCCACGACGAGCTCGAGCAGGAGCA	1752						
QY	211	GAAGCAAGCCAGGGACCCAGAGGAGGAGGCGCGGAGCAGCAGCTGAGCAGGCAAG	270						
DB	1753	GAAGCAAGCCAGGGACCCAGAGGAGGAGGCGCGGAGCAGCAGCTGAGCAGGCAAG	1812						
QY	271	ATGCTGAAATGTGAGGCCAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGT	330						
DB	1813	ATGCTGAAATGTGAGGCCAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGT	1872						
QY	331	GCAAGTGTGCCAGGCTTACTTGGGCGAGTGGAGCATGAGGACATCGACATCAGCAG	390						
DB	1873	GCAAGTGTGCCAGGCTTACTTGGGCGAGTGGAGCATGAGGACATCGACATCAGCAG	1932						
QY	391	ATGCCGTGGAGGACCTCACTGAGCGCCAGTGGGAGGACCTGACCCAGCAGTACTACTCC	450						
DB	1933	ATGCCGTGGAGGACCTCACTGAGCGCCAGTGGGAGGACCTGACCCAGCAGTACTACTCC	1992						
QY	451	TCGTTTCATGCGGATGTTTCATCTCAATTCAGAAATTAATTTTCGCAAGTCCAGGCTC	510						
DB	1993	TCGTTTCATGCGGATGTTTCATCTCAAATTCAGAAATTAATTTTCGCAAGTCCAGGCTC	2052						
QY	511	TGCTGAATAGAAATCAGTCTGTGAACCTTCAGACGACATTTGACGGGCTCCGGAAATCT	570						
DB	2053	TGCTGAATAGAAATCAGTCTGTGAACCTTCAGACGACATTTGACGGGCTCCGGAAATCT	2112						
QY	571	GGATTATAAGCCCGGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGG	630						
DB	2113	GGATTATAAGCCCGGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGG	2172						
QY	631	AGGAGATCTCTGGAGCTGGCAGCTGCAGACACACCTCTTCCAGGGACAACAAGTGGGTG	690						
DB	2173	AGGAGATCTCTGGAGCTGGCAGCTGCAGACACACCTCTTTCAGGGACAACAAGTGGGTG	2232						
QY	691	TCGAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT	750						
DB	2233	TCGAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT	2292						
QY	751	GGTTCTCTGTCGCGGACTGGAACCCCTGACCATCTGGTTCTACAAGGAGGTTACTTCC	810						
DB	2293	GGTTCTCTGTCGCGGACTGGAACCCCTGACCATCTGGTTCTACAAGGAGGTTACTTCC	2352						
QY	811	GGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACA	870						
DB	2353	GGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACA	2412						
QY	871	ACGCCGTCCAGAAGTACCTGAAGAATGATGTGGGCCGCGAGCCCCCTGCTGCCGCGCAC	930						

RESULT 4
US-10-635-977-12
; Sequence 12, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977

; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-12

Query Match 97.0%; Score 1782.8; DB 19; Length 3554;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1824; Conservative 0; Mismatches 2; Indels 30; Gaps 1;

Qy 1 TGGGAGCAGGCTGGGCCCCATCGGCTATGAGGCGGGAAGTGGGGCGGGTTGGGGAGCC 60
Db 1513 TGGGAGCAGGCTGGGCCCCATCGGCTATGAGGCGGGAAGTGGGGCGGGTTGGGGAGCC 1572

Qy 61 TCCGTGGCCCTGGGCCCCATCTCCAGTCCCAAGTCTCGCTGGACAGATAGGGCGAGGCT 120
Db 1573 TCCGTGGCCCTGGGCCCCATCTCCAGTCCCAAGTCTCGCTGGACAGATAGGGCGAGGCT 1632

Qy 121 GTGCTGTCTTTCA-----GAAGACTTCCGGCGCA 150
Db 1633 GTGCTGTCTTTCAAGAGTAGCAGCGTGGGAAGAGAGTTCCTGGAAAGACTTCCGGCGCA 1692

Qy 151 CCATGGCATCCAGCATCTCAAGTGGGTGGTACGCCACAGAGCTGCAGCAGGAGCAGCA 210
Db 1693 CCATGGCATCCAGCATCTCAAGTGGGTGGTACGCCACAGAGCTGCAGCAGGAGCAGCA 1752

Qy 211 GAAGCAAGCCAGGACACAGAGGAGGAGGCGCGGAGCAGCGACTTGAGCAGCAGGCAAG 270
Db 1753 GAAGCAAGCCAGGACACAGAGGAGGAGGCGCGGAGCAGCGACTTGAGCAGCAGGCAAG 1812

Qy 271 ATGCTGAATATGCTAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGT 330
Db 1813 ATGCTGAATATGCTAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGT 1872

Qy 331 GCAAGGTGTGCAGGCTTACTCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAG 390
Db 1873 GCAAGGTGTGCAGGCTTACTCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAG 1932

Qy 391 ATGCCGTGGAGGACCTCACTAGGCCCGAGTGGGAGGACCTGACCCAGCAGTACTCTCC 450
Db 1933 ATGCCGTGGAGGACCTCACTAGGCCCGAGTGGGAGGACCTGACCCAGCAGTACTCTCC 1992

Qy 451 TCGTTCAATGGGATGCTTTTCATCTCAATTCAGAAATTAATTTTCGAGTGCAGGCTC 510
Db 1993 TCGTTCAATGGGATGCTTTTCATCTCAATTCAGAAATTAATTTTCGAGTGCAGGCTC 2052

Qy 511 TGCTGAATAGAAATCAAGCTCTGTGAACCTCAGACGGACATTTGAGCGGCTCCGGAAATCT 570
Db 2053 TGCTGAATAGAAATCAAGCTCTGTGAACCTCAGACGGACATTTGAGCGGCTCCGGAAATCT 2112

Qy 571 GGATTATAAGCCCGGCCCAAGTCCCGGGGCGGAGACATAGTGTGATGAGACCGGTGG 630
Db 2113 GGATTATAAGCCCGGCCCAAGTCCCGGGGCGGAGACATAGTGTGATGAGACCGGTGG 2172

Qy 631 AGGAGATCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAAACAAGTGGGTGG 690
Db 2173 AGGAGATCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAAACAAGTGGGTGG 2232

Qy 691 TCCAGAGATACATCGAGACCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 750
Db 2233 TCCAGAGATACATCGAGACCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 2292

Qy 751 GGTTCCTCGTCAGGACTGGAACCCCTGACCATCTGGTTCTACAGGAGAGTACTTGC 810
Db 2293 GGTTCCTCGTCAGGACTGGAACCCCTGACCATCTGGTTCTACAGGAGAGTACTTGC 2352

RESULT 5
US-10-615-659-11
; Sequence 11, Application US/10615659
; Publication No. US20040157234A1

Qy 811 GGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGGCATCCACTGTGCAACA 870
Db 2353 GGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGGCATCCACTGTGCAACA 2412

Qy 871 ACGCGGTCCAGAAATACCTGAAGAATGATGTGGCGCGCAGCCCCCTCTGCTGCCCGCACACA 930
Db 2413 ACGCGGTCCAGAAATACCTGAAGAATGATGTGGCGCGCAGCCCCCTCTGCTGCCCGCACACA 2472

Qy 931 ACATGTGGACACGACACAGGTTCCAGAGTACCTTGACGCCACAGGGCGGTGGCGCCGTGT 990
Db 2473 ACATGTGGACACGACACAGGTTCCAGAGTACCTTGACGCCACAGGGCGGTGGCGCCGTGT 2532

Qy 991 GGGCGACGCTCATCTACCCGCTCCATGAAGAAGCCATCGCCCGCAGCCATGAAGGTGGCCC 1050
Db 2533 GGGCGACGCTCATCTACCCGCTCCATGAAGAAGCCATCGCCCGCAGCCATGAAGGTGGCCC 2592

Qy 1051 AGGACACGTGGAGCTTCGCAAGAACAGCTTTTCAGAGCTCTACGGGGCTGACTTTCGTCCTTG 1110
Db 2593 AGGACACGTGGAGCTTCGCAAGAACAGCTTTTCAGAGCTCTACGGGGCTGACTTTCGTCCTTG 2652

Qy 1111 GAGGGGACTTCAGGGCTTGTGCTGATCGAGATCAATTCAGCCCGCACCATGCAACCCGTCCA 1170
Db 2653 GAGGGGACTTCAGGGCTTGTGCTGATCGAGATCAATTCAGCCCGCACCATGCAACCCGTCCA 2712

Qy 1171 CGCGGCTCAGCGGCCAGCTGTGTGCAAGGTGCGAGGAGCACCATCAAGGTGGCCGCTGG 1230
Db 2713 CGCGGCTCAGCGGCCAGCTGTGTGCAAGGTGCGAGGAGCACCATCAAGGTGGCCGCTGG 2772

Qy 1231 ACCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGAGCGCGGTGGTGGAGCGCG 1290
Db 2773 ACCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGAGCGCGGTGGTGGAGCGCG 2832

Qy 1291 CCCCATTCAGCGGTTCGACCTCTGCGTGGCGGGCTCAGTGTGAGGAGAGCCAGGAGGC 1350
Db 2833 CCCCATTCAGCGGTTCGACCTCTGCGTGGCGGGCTCAGTGTGAGGAGAGCCAGGAGGC 2892

Qy 1351 AGGTGTGCGCGCTCTGCACCTCAAGGCTCGCGCTCGCTGTGGAGCGCGCAGCGCTGA 1410
Db 2893 AGGTGTGCGCGCTCTGCACCTCAAGGCTCGCGCTCGCTGTGGAGCGCGCAGCGCTGA 2952

Qy 1411 AGGACGGGGCCCTCGGCCATGCCAGACCTCGCCAGGGAGCCCCCATCACAGCTCTCC 1470
Db 2953 AGGACGGGGCCCTCGGCCATGCCAGACCTCGCCAGGGAGCCCCCATCACAGCTCTCC 3012

Qy 1471 AGCGGACTTTGGGACTGAAGGAGAGAGGGGTCTCCCTCGGCTTGTCTGGCACCCCTTAA 1530
Db 3013 AGCGGACTTTGGGACTGAAGGAGAGAGGGGTCTCCCTCGGCTTGTCTGGCACCCCTTAA 3072

Qy 1531 GGGGGGACCGGAGGGGTGGAGCCGACAGCCCGCAGCCCGCACCAAGCTGCTGGGAGG 1590
Db 3073 GGGGGGACCGGAGGGGTGGAGCCGACAGCCCGCAGCCCGCACCAAGCTGCTGGGAGG 3132

Qy 1591 TGGAGTCTCCCGGCTCGCCCTGTGCGCACGTGAGCAGTCAAGGCCCAAAACACACCGGTGTC 1650
Db 3133 TGGAGTCTCCCGGCTCGCCCTGTGCGCACGTGAGCAGTCAAGGCCCAAAACACACCGGTGTC 3192

Qy 1651 CCGTAGCCACGCGCCCAAAAGCTGGGATCCAAACAGCTAAATGCGCACCCCGCTGGAGC 1710
Db 3193 CCGTAGCCACGCGCCCAAAAGCTGGGATCCAAACAGCTAAATGCGCACCCCGCTGGAGC 3252

Qy 1711 CTGTGCTGGGGGCTTGAAGACAGAGGGCGCGCTCGCTCCCGCCCGGAGGAGAAAG 1770
Db 3253 CTGTGCTGGGGGCTTGAAGACAGAGGGCGCGCTCGCTCCCGCCCGGAGGAGAAAG 3312

Qy 1771 GTTCATGACAGGCTCAGATTCTCTGACAGGAGGTACAGGTTTCAGGCACTCTCCC 1826
Db 3313 GTTCATGACAGGCTCAGATTCTCTGACAGGAGGTACAGGTTTCAGGCACTCTCCC 3368

Qy 1741 GCGCGCTGCGTCCGCGCCGAGGAAAGGTTTCATGACAGCTCAGATTCTCTGACGCA 1800
Db |||||||
Qy 1674 GCGCGCTGCGTCCGCGCCGAGGAAAGGTTTCATGACAGCTCAGATTCTCTGACGCA 1733
Db |||||||
Qy 1801 GGAGTACAGGTTGACGACCTCTCC 1826
Db |||||||
Qy 1734 GGAGTACAGGTTGACGACCTCTCC 1759
Db |||||||
RESULT 8
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match 91.3%; Score 1677.2; DB 19; Length 1939;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1756; Conservative 0; Mismatches 3; Indels 67; Gaps 1;

Qy 1 TGGGAGCAGGCTGGGCCCCATCGGCTATGAGGGCGGGAAGTGGGGCGGGTTGGGGAGCC 60
Db |||||||
Qy 61 TCCGTGGCCCTGGCCCATCTCCAGTCCCAGTCTCGCTCGGACAGATAGGGCGAGGCT 120
Db |||||||
Qy 121 GTGCTGTCTTTTCAGAAAGCTTCGGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG 180
Db |||||||
Qy 181 TCAGCCACAGAGCTGAGCAGGAGCAGAGCAGGAGCAGGAGCAGGAGGAGG 240
Db |||||||
Qy 241 CCGGGAGCAGGCTTCAGCAGCAGGAGCAGATGCTGAAATGCTGAGGCAAAAGCTCAGGG 300
Db |||||||
Qy 301 GCCTCCCGGGGAGCTTTGTGACATCGCGTGCAGAGGTGTCAGGCGCTACCTGGGGCAGC 360
Db |||||||
Qy 361 TGGAGCATGAGGACATCGACATCGGTGCAAGGTGTCAGGCGCTACCTGGGGCAGC 420
Db |||||||
Qy 421 GGGAGGACCTGACCCAGCAGTACTCTCCCTCGTTTCATGGGATGCTTTCATCTCCAATT 480
Db |||||||
Qy 481 CAAGAAATTACTTTTCAGTGGCAGGCTCTGTGAAATAGAAATCACGCTCTGTGAAACCTTC 540
Db |||||||

Qy 541 AGACGGACATTGACGGGCTCCGGAAACATCTGGATTATATAAGCCGCGGCCCAAGTCCCGGG 600
Db |||||||
Qy 601 GCGGACACATAGTGTGCATGGACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACC 660
Db |||||||
Qy 661 ACCTCTTTTCAGGGAACAAGTGGGTGTCAGAAAGTACATCGAGACGCGCGTGTCTCA 720
Db |||||||
Qy 721 TCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTCGA 780
Db |||||||
Qy 781 CCATCTGTTTCTACAAGGAGGATTTACTTTGCGGTTTCTCAACTCAGCGCTTCTCCCTGGACA 840
Db |||||||
Qy 841 AGCTGACAGCGCCATTCACCTGTGCAACAAACGCGTTCAGAAAGTACCTGAAAGTATG 900
Db |||||||
Qy 901 TGGGCGCAGCCCTGCTGCGCCGACACAAATGTGAGCCAGCAGGTTCCAGGAGT 960
Db |||||||
Qy 961 ACCTGACGCGCAGGCGCGTGGCGCGTGTGGGGCAGCGTCATCTACCCGTCATGAAGA 1020
Db |||||||
Qy 1021 AGGCCATCGCCCAAGGTCAGGTCGCGGTCAGAGCAGCTTCGCAAGAACAGCT 1080
Db |||||||
Qy 1081 TTGAGCTCTACGCGGCTGACTTCGTCTTGGGAGGAGTTCAGGCGCTGCTGATCCAGA 1140
Db |||||||
Qy 1141 TCAATTCAGGCCCCACCATGCACCCGTCACGCGCGGTACGCGCCAGCTGTGTGCACAGG 1200
Db |||||||
Qy 1201 TGACGAGGACACCATCAAGGTGCGCGTGAACGCGAGCTGTGACATCGGCAACTTCGAGC 1260
Db |||||||
Qy 1261 TCCTGTGGAGGACCGGCTGTTGAGCGCGCCCAATTCAGCGGGTCCGAGCTCTGCGTGG 1320
Db |||||||
Qy 1321 CGGCGCTCAGTGTGAGGAGCAGGAGCAGGTCGTGTCGCCGTCTGCAACCTCAAGGCT 1380
Db |||||||
Qy 1381 CGGCGCTCAGTGTGAGGAGCAGGAGCAGGTCGTGTCGCCGTCTGCAACCTCAAGGCT 1440
Db |||||||
Qy 1441 CTGCGGAGGACCCCATCATCAGCTCTCCAGCGGACTTTGGGACTGGAAGGAGAGG 1500
Db |||||||
Qy 1501 GGCTCCCGCTTGGCTTGGGCAACCTTAAGGGGGGCGAGCGGCTGGAGCCGCGC 1560
Db |||||||
Qy 1561 AGCCCAACCCGACCAAGAGTGTGGGAGGTGAGCTCCCGGCTGCGCTGTGCGCACG 1620
Db |||||||
Qy 1621 TGGACAGTACGGCCCCCAACACCGGTTGTCCCGTAGCCAGCGCGCCCAAGAGCTGGGATC 1680

Db 1542 GAGCACCCTGGAGCTGTCTGGGAGCTGAGACAGAGGGCGCTGGCTCCG 1601
Qy 1755 CCGCCCGGAGAAAGTTTCATGACAGCGTCAGATTCTTGCAGCAGGAGTACAGGTTGC 1814
Db 1602 CCGCCCGGAGAAAGTTTCATGACAGCGTCAGATTCTTGCAGCAGGAGTACAGGTTGC 1661
Qy 1815 AGCACTCTCCC 1826
Db 1662 AGCACTCTCCC 1673

RESULT 10
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10

Query Match 85.7%; Score 1574.8; DB 19; Length 1859;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1643; Conservative 0; Mismatches 2; Indels 47; Gaps 2;

Qy 135 GAAGACTTCGGCGCACCAATGGCATCAGCATCCTCAAGTGGGTGTCAGCCACCAAGAGC 194
Db 29 GAAGACTTCGGCGCACCAATGGCATCAGCATCCTCAAGTGGGTGTCAGCCACCAAGAGC 88
Qy 195 TGCAGCAGGAGCAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 254
Db 89 TGCAGCAGGAGCAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 148
Qy 255 CTGAGCAGCAGGCAAGATGCTGAAATGCTGAGCAGCAAGCTCAGGGGCTCCCGGGCAG 314
Db 149 CTGAGCAGCAGGCAAGATGCTGAAATGCTGAGCAGCAAGCTCAGGGGCTCCCGGGCAG 208
Qy 315 CTTGTGGACATCGCTGCAAGGTGTGCGCAGGCCCTACCTGGGGCAGCTGGAGCATGAGGAC 374
Db 209 CTTGTGGACATCGCTGCAAGGTGTGCGCAGGCCCTACCTGGGGCAGCTGGAGCATGAGGAC 268
Qy 375 ATGACACGTCAGCAGATGCGGTGAGGAGCCTCATGTAGGCCAGTGGAGGAGCCTTGACC 434
Db 269 ATGACACGTCAGCAGATGCGGTGAGGAGCCTCATGTAGGCCAGTGGAGGAGCCTTGACC 328
Qy 435 CAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAAGAAATTACTTT 494
Db 329 CAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAAGAAATTACTTT 351
Qy 495 TCGCAGTGGCAGGCTCTGCTGAATAGAAATCACGTCGTGTGAACCTCTCAGACGGACATTGAC 554
Db 352 -----TGCCAGGCTCTGCTGAATAGAAATCACGTCGTGTGAACCTCTCAGACGGACATTGAC 405
Qy 555 GGCGCTCGGAAACATCTGGATATTAAGCCCGCGGCCAAGTCCCGGGGCCGAGACATAGTG 614
Db 406 GGCGCTCGGAAACATCTGGATATTAAGCCCGCGGCCAAGTCCCGGGGCCGAGACATAGTG 465
Qy 615 TGCATGACCGCTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGG 674

Db 466 TGCATGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCTCTTTCCAGG 525
Qy 675 GACAAACAAGTGGGTGTTCAGAAAGTACATCGAAGCGCGCTGTCTCATCTGTGACACCAAG 734
Db 526 GACAAACAAGTGGGTGTTCAGAAAGTACATCGAAGCGCGCTGTCTCATCTGTGACACCAAG 585
Qy 735 TTCGACATCAGACAGTGGTTCTTCGTCACGAGCTTGGAAACCCCTTGACCATCTGGTTCTAC 794
Db 586 TTCGACATCAGACAGTGGTTCTTCGTCACGAGCTTGGAAACCCCTTGACCATCTGGTTCTAC 645
Qy 795 AAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCC 854
Db 646 AAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCC 705
Qy 855 ATCCACCTGTGCAACAACCGCGTTCAGAAAGTACATGAAAGTATGATGTGGGCGCGAGCGCC 914
Db 706 ATCCACCTGTGCAACAACCGCGTTCAGAAAGTACATGAAAGTATGATGTGGGCGCGAGCGCC 765
Qy 915 CTGCTGCCCGCACACAACATGTGACAGCAGCAGGTTCCAGAGTACCTTGCAGCGCAG 974
Db 766 CTGCTGCCCGCACACAACATGTGACAGCAGCAGGTTCCAGAGTACCTTGCAGCGCAG 825
Qy 975 GGCGGTGGCGCGTGTGGGCGAGCGTCATCTACCCGTCCTCAATGAAGAGGCGCATCGCCAC 1034
Db 826 GGCGGTGGCGCGTGTGGGCGAGCGTCATCTACCCGTCCTCAATGAAGAGGCGCATCGCCAC 885
Qy 1035 GCCATGAAGTGGCCCGCAGAGCAGCAGTGGAGCTCGCAAGAACAGCTTTGAGTCTACGGG 1094
Db 886 GCCATGAAGTGGCCCGCAGAGCAGCAGTGGAGCTCGCAAGAACAGCTTTGAGTCTACGGG 945
Qy 1095 GCTGACTTCTGCTTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTTCCAGCGCC 1154
Db 946 GCTGACTTCTGCTTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTTCCAGCGCC 1005
Qy 1155 ACCATGACACCGCTCCACGCGGTTCAGGCGCCAGCTGTGTGCACAGGTGCGAGGAGACAC 1214
Db 1006 ACCATGACACCGCTCCACGCGGTTCAGGCGCCAGCTGTGTGCACAGGTGCGAGGAGACAC 1065
Qy 1215 ATCAAGGTGGCGCGTGCAGCGCAGCTGTGACATCGGCAACTTCGAGTCTCTGTGAGGCGAG 1274
Db 1066 ATCAAGGTGGCGCGTGCAGCGCAGCTGTGACATCGGCAACTTCGAGTCTCTGTGAGGCGAG 1121
Qy 1275 CGGCTGTTGAGCGCGCCCATTTTCAGGCGCTCGACCTCTGCTGGCGGCGCTCAGGTG 1334
Db 1122 CGGCTGTTGAGCGCGCCCATTTTCAGGCGCTCGACCTCTGCTGGCGGCGCTCAGGTG 1181
Qy 1335 AGGAGACCGCAGGAGGAGTGTGCTGCGCTCTGCAACCTCAAGGCTCTCGGCTCTGCTTTG 1394
Db 1182 AGGAGACCGCAGGAGGAGTGTGCTGCGCTCTGCAACCTCAAGGCTCTCGGCTCTGCTTTG 1241
Qy 1395 GACGCGAGCGCTGAAGGACAGCGGCGCTTCGGCCATGCGACCTGCGCCAGGAGACCC 1454
Db 1242 GACGCGAGCGCTGAAGGACAGCGGCGCTTCGGCCATGCGACCTGCGCCAGGAGACCC 1301
Qy 1455 CACATCACAGCTCTCAGCGGAGCTTTGGGACTTGAAGAGAGAGGAGGCTCCCGCTGGCC 1514
Db 1302 CCATCACAGCTCTCAGCGGAGCTTTGGGACTTGAAGAGAGAGGAGGCTCCCGCTGGCC 1361
Qy 1515 TTGCTGGCAGCCCTTAAAGGGGGCAGCGAGCGGTGGAGCCGACAGCCCAAGCCGAGCC 1574
Db 1362 TTGCTGGCAGCCCTTAAAGGGGGCAGCGAGCGGTGGAGCCGACAGCCCAAGCCGAGCC 1421
Qy 1575 AAAAGTGTGGGAGGTGGAGTCCCGGCTGCGGCTGCGCCCTGTGCGCAGTGGAGTCAAGTCCG 1634
Db 1422 AAAAGTGTGGGAGGTGGAGTCCCGGCTGCGGCTGCGCCCTGTGCGCAGTGGAGTCAAGTCCG 1481
Qy 1635 CAAAAACCGGCTCTCCCGTAGCCCGCCGCAAAAGCTGGGATCCAAACCCAGCTAAAT 1694
Db 1482 CAAAAACCGGCTCTCCCGTAGCCCGCCGCAAAAGCTGGGATCCAAACCCAGCTAAAT 1541
Qy 1695 GCGCACCGCTGGAGGCTGTGCTGGGGGCTTGAAGACAGCAGAGGCGGCTGCGCTCCG 1754
Db 1542 GAGCACCGCTGGAGGCTGTGCTGGGAGCTTGAAGACAGCAGAGGCGGCTGCGCTCCG 1601

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-635-977-3

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e-189; Length 726;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

Qy 549 ATTGACGGGCTCCGGAAACATCTGGAATTATAAAGCCCGCGGCCAAGTCCCGGGGCCGAGAC 608
|||
Db 1 ATTGACGGGCTCCGGAAACATCTGGAATTATAAAGCCCGCGGCCAAGTCCCGGGGCCGAGAC 60

Qy 609 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACACACCCCTCTT 668
|||
Db 61 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGCGAGCTGCAGACACACCCCTCTT 120

Qy 669 TCCAGGACAAACAGTGGGTGGTCCAGAAGTACATCGAGACGCGGTGCTCATCTGTGAC 728
|||
Db 121 TCCAGGACAAACAGTGGGTGGTCCAGAAGTACATCGAGACGCGGTGCTCATCTGTGAC 180

Qy 729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACCTGGAACCCCTGACCATCTGG 788
|||
Db 181 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACCTGGAACCCCTGACCATCTGG 240

Qy 789 TTCTTCAAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 848
|||
Db 241 TTCTTCAAGGAGAGTTACTTTGCGGTTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 300

Qy 849 AGGCCCATCCACTGTGCACAACAGCGGTCAGAGTACTGCTGAGANTGATGTGGCGCCG 908
|||
Db 301 AGGCCCATCCACTGTGTCAACAACAGCGGTCAGAGTACTGGAAGAAATGATGTGGCGCCG 360

Qy 909 AGCCCCCTGTCGCCGCGACACAACATGTGGACGAGCACAGGTTTCCAGGAGTACTCTGAG 968
|||

```

Db 361 AGCCCCCTGCTGCCCGCACACATGTGACACAGACACAGGTTCACAGAGTACCTGCAG 420
Qy 969 CGCAGAGCCCGTGGCGCGTGTGGGCGAGCGTCACTTACCGCTCCATGAAGAAGGCCATC 1028
Db 421 CGCAGAGCCCGTGGCGCGTGTGGGCGAGCGTCACTTACCGCTCCATGAAGAAGGCCATC 480
Qy 1029 GCCCAGCCATGAAGTGGCCCGAGGACCACTGGAGCCTCGCAAGACAGCTTTGAGCTC 1088
Db 481 GCCCAGCCATGAAGTGGCCCGAGGACCACTGGAGCCTCGCAAGACAGCTTTGAGCTC 540
Qy 1089 TAGCGGCGCTGACTTTCCTTGGGAGGAGCTTCAGGCGCTCGGTGATGCGAGATCAATTCC 1148
Db 541 TAGCGGCGCTGACTTTCCTTGGGAGGAGCTTCAGGCGCTCGGTGATGCGAGATCAATTCC 600
Qy 1149 AGCCCCACCATGACCCCGTCCAGCCCGTCAAGCCCGTCAAGCCCGTCAAGTGTGACAGGAG 1208
Db 601 AGCCCCACCATGACCCCGTCCAGCCCGTCAAGCCCGTCAAGCCCGTCAAGTGTGACAGGAG 660
Qy 1209 GACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCTGTGG 1268
Db 661 GACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCTGTGG 720
Qy 1269 AGGCAG 1274
Db 721 AGGCAG 726
RESULT 13
US-10-723-860-631/c
; Sequence 631, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma. Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 631
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-631
Query Match 27.3%; Score 501; DB 20; Length 101270;
Best Local Similarity 100.0%; Pred. No. 3.2e-127;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1272 CAGCCGCTGTTAGCCGCCGCCATTCAGCGGGTCCGACCTCTGCTGGCGGGCGTCACT 1331
Db 43637 CAGCCGCTGTTAGCCGCCGCCATTCAGCGGGTCCGACCTCTGCTGGCGGGCGTCACT 43578
Qy 1332 GTGAGGAGAGCCAGGAGGAGGTGCTGCCCGTCTGCAACCTCAAGGCTCGGCTCGCTG 1391
Db 43577 GTGAGGAGAGCCAGGAGGAGGTGCTGCCCGTCTGCAACCTCAAGGCTCGGCTCGCTG 43518
Qy 1392 TTGAGCGCGCAGCGCTGGAAGGACCGGGGCCCTCGGCCATGCGACACCTGCCAGGGA 1451
Db 43517 TTGAGCGCGCAGCGCTGGAAGGACCGGGGCCCTCGGCCATGCGACACCTGCCAGGGA 43458
Qy 1452 CCCCCATCACAGCTCTCCAGCGGAGTCTGGGACTGAAGAGAGAGAGGGGCTCCCGCTG 1511
Db 43457 CCCCCATCACAGCTCTCCAGCGGAGTCTGGGACTGAAGAGAGAGAGGGGCTCCCGCTG 43398
Qy 1512 GCCTTGTGACCCCTTAAGGGGGGCGAGCGAGCGGTGAGCGCGGCGGACGACAGCCCGC 1571

Db 43397 GCCTTGTGACCCCTTAAGGGGGGCGAGCGAGCGGTGAGCGCGCACAGCCACCCCGC 43338
Qy 1572 ACCAAAGCTGCTGGGAAGGTGGAGCTCCCGGCCCTGCCCCCTGTGCCACGCTGGACAGTCAG 1631
Db 43337 ACCAAAGCTGCTGGGAAGGTGGAGCTCCCGGCCCTGCCCCCTGTGCCACGCTGGACAGTCAG 43278
Qy 1632 GCCCCAAACACACCGTGTCCCCGTAGCCCGCCGCAAAAGCTGGGATCCAAACCAAGCTA 1691
Db 43277 GCCCCAAACACCGTGTCCCCGTAGCCCGCCGCAAAAGCTGGGATCCAAACCAAGCTA 43218
Qy 1692 AATGCGCACCCGCTGGAGCCTGTGCTGCGGGGCTTGAAGACAGCAGAGGCGCGCTCGCT 1751
Db 43217 AATGCGCACCCGCTGGAGCCTGTGCTGCGGGGCTTGAAGACAGCAGAGGCGCGCTCGCT 43158
Qy 1752 CGCCCGCCCGGAGGAAAAGT 1772
Db 43157 CGCCCGCCCGGAGGAAAAGT 43137
RESULT 14
US-10-275-595A-39
; Sequence 39, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 2380
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CB1
US-10-275-595A-39
Query Match 23.5%; Score 432.8; DB 18; Length 2380;
Best Local Similarity 59.1%; Pred. No. 1e-108;
Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;
Qy 260 CAGCAGGAGAGTCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCTCCCGGGGAGCTTG 318
Db 361 CAGGACACAGCAGCGCCCAAGAAAACCCAGTGTGTTGCTGCCAGAGTTTG 420
Qy 319 TGGACATCGCGTCAAGGTGTGCCAGGCTTACCTGGGCGAGCTGGAGCATGAGGACATCG 378

```
Db 421 TGGATGAAGCTCTGTGTGGTGGAGAGTACCTTAGCAACTTTGGCCCATGTGGAATCG 480
Qy 379 ACACGTGACAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGACCTGACCCAGC 438
Db 481 ACAAGGACCTGGAGGCCCGCTGTACTCTACCCCGAGGGCTGGTCCCTCTTCCCTCAGC 540
Qy 439 AGTACTACTCCCTCGTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATATCTTTTCG 498
Db 541 GCTACTTACCAAGTGTTCACGAAGGGGCGAGAACTCAGGCACCTCGACACTCAGGTCAGC 600
Qy 499 AGTGCCAGGCTCTGCTGAATAGNATCACGTCGTGTGAACCTCAGACGGACATTCAGGGC 558
Db 601 GCTGTGAGGACATCTCTGCGACAGCTCGAGCGCGCTGTATCCCGACATAGACATGGAAGGG 660
Qy 559 TCGGAACATCTGGATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGCA 618
Db 661 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGGCATCATGTGCA 720
Qy 619 TGGACCGTGTGAGGAGATCTTGGAGCTGGCAGCTGCGACGACCAACCTCTTTCCAGGGA 678
Db 721 TGGACCACCTGGAGGAGATGCTGAGCTGCTGAACCGCAACCCCGTGTGTGATGAAGGACG 780
Qy 679 ACAAGTGGTGTCCAGAGTATCATCGAGACCGCTGCTCATCTGTGACACCAAGTTTCG 738
Db 781 GCAAGTGGTGTGCGAAGATATATTGAGCGCGCCCTCTCATCTTTTGGCACCAAGTTTG 840
Qy 739 ACATCAGACAGTGGTTTCCTCGTCACGAGCTGGAACCCCTCGACCATCTGGTTCTCAAGG 798
Db 841 ACCTCAGACAGTGGTTTCCTGTTAACTGACTGGNAACCACTTACCGTGTGTTTACCGCG 900
Qy 799 AGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCTTGACAAAGCTGGACAGCGCATCC 858
Db 901 ACAGCTATATCCGCTTTTCCAGCGCAGCCCTTCTCCTGGAAGAACCTGGACAACTCAGTGC 960
Qy 859 ACTGTGTGAACACGCGTCCAGAGTACTGAGATCATGAGATCATGTGGCCCGCAGCCCTTCG 918
Db 961 ACTGTGTGAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCGCATCGGCATCCACTGC 1020
Qy 919 TGCCCGACACAACATGTGGACAGACCAACAGGTTCCAGAGTACCTGACGCGCCAGGGCC 978
Db 1021 TTCCGCGCAGAACATGTGCTTAGCCAGAGTTTCAGGCCCACTGACGAGATGGGTG 1080
Qy 979 GTGGCCCGCTGTGGGCGACGCTCATCTACCGTCCATGAAGAGGCCATCGGCCACGCCA 1038
Db 1081 CCCCMAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGATGTGTGATCCACGCCAC 1140
Qy 1039 TGAAGTGGCCCGACGACAGCTGGAGCTTCGCAAGAACAGCTTGAGCTCTACGGGGCTG 1098
Db 1141 TTCAGACCTCCACAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTATGGCGCTG 1200
Qy 1099 ACTTCGCTCTGGGAGGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCCCCACC 1158
Db 1201 ACTTCGTTTCGGGAGGAGTTCAGCCCTGCTGATTCAGATCAACGCGCAGGCCACCA 1260
Qy 1159 TGACCCCGTCCACGCGGTCAAGGCCACGCTGTGTGCAAGGTGACAGGAGGACACCATCA 1218
Db 1261 TGGCACCTCCACAGCAGTCACTGCCCCGCTGTGCTGGCGTGCAGCTGCAACCTGC 1320
Qy 1219 AGTGGCC-----GTGGACCGAGCTGTGACATCGCACTTCGAGCTCTCTGT 1266
Db 1321 GCGTGTGATATTACCGGATGTGTGACCGCAACTGTGACACAGAGGCTTTTGAGCTCATCT 1380
Qy 1267 GGAGGCGAGCGGTGTGTGAGCGCGCCCACTTCAGCGGGTCCGACCTCTGCGTGGCGGGCG 1326
Db 1381 ATAAAGCAGCTCTGTGTGAGGTGCCTCAATATGTGGGCATCCGGCTCTCTGTAGAGGCT 1440
Qy 1327 TCAGTGTGAGGAGAGCCAGAGGACAGGTGTGCCCTGTGTGCAACTCAAGGCGCTCGGCT 1386
Db 1441 TCACCATCAAGAAGCCCATGGCGATGTGTATCGGCGGATGGGGTCCGCCACGAGCTCC 1500
Qy 1387 CGCTGTGTGACGCGCAGCCGCTGAAGGACAGGGGCCCCCTCGGCCATGCC-AGACCCCTGCC 1445
Db 1501 CTCTGCTGACCCAGCAGGCTCTTGGGGAAGGCAAGGACTCGGGGATTCCTTACCCACAGGT 1560
```

```
Qy 1446 CAGGAGCCCCATCACAGCTCTCAGCGGGACTTGGGACTGAAAGAGAGAGGGGCTC 1505
Db 1561 CAGCTTCTAGGAAAGGCACTGGGGCCAGGAGCTGGGGCACAGTGAAGACCAGTCTCCA 1620
Qy 1506 CCCCTGGCTTCTGTCACCCCTTAAGGGGGGAGCCGAGAGCGGTGGAGCCGACAGCCC 1565
Db 1621 CTGCGACCACTTTCAGCCCCCGGAAAGGGAAGGCAAGGCAAAAGGGCCACAGCCC 1680
Qy 1566 ACCCGCACCACCAAGCTGCTGGGAAGGTGGAGCTCCCGGC 1603
Db 1681 TGGTCTGCCCAATCTCTGGGAGTGGGATGCCCCGAGC 1718
```

RESULT 15

```
US-10-029-386-24894/c
; Sequence 24894, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AROMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24894
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022327.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.35
; OTHER INFORMATION: SWISSPROT HIT: P38160, EVALUUE 5.00e-04
; OTHER INFORMATION: EST HUMAN HIT: BE82753.1, EVALUUE 3.00e-06
; OTHER INFORMATION: NT HIT: gill4726145, EVALUUE 8.00e-09
US-10-029-386-24894
```

Query Match 23.2%; Score 427; DB 16; Length 490;

Best Local Similarity 100.0%; Pred. No. 3e-107;

Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 848 CAGCGCCATCCACCTGTGCAACAACGCGTCCAGAACTACCTGAAGAAATGATGTGGCCG 907
Db 427 CAGCGCCATCCACCTGTGCAACAACGCGTCCAGAACTACCTGAAGAAATGATGTGGCCG 368
Qy 908 CAGCCCCCTGTGCCCGCACACAACATGTGGACCAACAGGTTCCAGGAGTACTGCA 967
Db 367 CAGCCCCCTGTGCCCGCACACAACATGTGGACCAACAGGTTCCAGGAGTACTGCA 308
Qy 968 GCGCCAGGCGCGTGGCGCGGTGGGGCAGCGTCACTACCCGCTCCATGAAGAGCCAT 1027
Db 307 GCGCCAGGCGCGTGGCGCGGTGGGGCAGCGTCACTACCCGCTCCATGAAGAGCCAT 248
Qy 1028 CCGCCACGCGCATGAAGTGGCCAGACCACTGGAGCTTCGCAAGAACAGCTTTGAGCT 1087
Db 247 CCGCCACGCGCATGAAGTGGCCAGACCACTGGAGCTTCGCAAGAACAGCTTTGAGCT 188
Qy 1088 CTACGGGCTGACTTTCGTCTTTGGAGGGAATTCAGGGCCCTGGCTGATCGAGATCAATTC 1147
Db 187 CTACGGGCTGACTTTCGTCTTTGGAGGGAATTCAGGGCCCTGGCTGATCGAGATCAATTC 128
Qy 1148 CAGCCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1207
Db 127 CAGCCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 69
Qy 1208 GGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGATCGGCAACTTCGAGCTCTCTGTG 1267
Db 67 GGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGATCGGCAACTTCGAGCTCTCTGTG 8
```

Qy 1268 GAGGAG 1274
| | | | |
Db 7 GAGGAG 1

Search completed: September 24, 2005, 16:48:37
Job time : 1423.93 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 20:40:04 ; Search time 335.858 Seconds
(without alignments) 8954.609 Million cell updates/sec

Title: US-10-615-659-1
Perfect score: 1838
Sequence: 1 tgggagcaggcctgggccc.....actctcccaaggcgcaattc 1838

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 24055568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum match 100%
Listing first 45 summaries

```
Database : Issued Patents_NA.*
1 : /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2 : /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3 : /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4 : /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5 : /cgn2_6/ptodata/1/ina/PTCTUS_COMB.seq.*
6 : /cgn2_6/ptodata/1/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	421.4	22.9	3001	4	US-09-620-312D-494	Sequence 494, Appl
2	56	3.0	909	4	US-09-252-991A-5992	Sequence 592, Ap
3	56	3.0	1305	4	US-09-252-991A-5951	Sequence 591, Ap
4	52.6	2.9	1272	4	US-09-252-991A-3471	Sequence 2471, Ap
5	52.6	2.9	1449	4	US-09-252-991A-3200	Sequence 2200, Ap
6	52.6	2.9	2790	4	US-09-252-991A-2632	Sequence 2632, Ap
7	50.4	2.7	1524	4	US-09-807-897-3	Sequence 3, Appli
8	50.2	2.7	9039	4	US-09-902-540-983	Sequence 983, App
9	48.6	2.6	2277	1	US-08-676-967-5	Sequence 5, Appli
10	48.6	2.6	2277	1	US-08-676-974-5	Sequence 5, Appli
11	48.6	2.6	2272	2	US-09-098-487-5	Sequence 5, Appli
12	48.2	2.6	2004	1	US-08-471-033-18	Sequence 18, Appl
13	48.2	2.6	2004	2	US-08-471-044-18	Sequence 18, Appl
14	48.2	2.6	2004	2	US-08-463-483A-18	Sequence 18, Appl
15	48.2	2.6	2004	2	US-08-471-046A-18	Sequence 18, Appl
16	48.2	2.6	2004	2	US-08-470-566B-18	Sequence 18, Appl
17	48.2	2.6	2004	2	US-08-469-334-18	Sequence 18, Appl
18	48.2	2.6	2004	3	US-09-300-529-18	Sequence 18, Appl
19	48.2	2.6	2555	1	US-08-471-033-17	Sequence 17, Appl
20	48.2	2.6	2555	2	US-08-471-044-17	Sequence 17, Appl
21	48.2	2.6	2555	2	US-08-463-483A-17	Sequence 17, Appl
22	48.2	2.6	2555	2	US-08-471-046A-17	Sequence 17, Appl
23	48.2	2.6	2555	2	US-08-470-566B-17	Sequence 17, Appl
24	48.2	2.6	2555	2	US-08-469-334-17	Sequence 17, Appl
25	48.2	2.6	2555	3	US-09-300-529-17	Sequence 17, Appl
26	48	2.6	5314	4	US-09-620-312D-138	Sequence 138, App
27	47.6	2.6	9927	4	US-09-902-540-9188	Sequence 9188, Ap

C	28	47.6	2.6	4884	4	US-09-252-991A-12126	Sequence 12126, App
	29	47.6	2.6	4884	4	US-09-252-991A-12126	Sequence 12126, App
	30	46.4	2.5	4600	4	US-09-702-705A-12292	Sequence 12292, App
	31	46.4	2.5	4600	4	US-09-702-705A-12292	Sequence 12292, App
	32	46.4	2.5	4600	4	US-09-736-457-1797	Sequence 1797, App
	33	46.4	2.5	4600	4	US-09-671-325-1797	Sequence 1797, App
	34	46	2.5	19455	4	US-09-902-540-1147	Sequence 1147, App
	35	46	2.5	1218	4	US-09-902-540-5379	Sequence 5379, App
	36	46	2.5	2874	4	US-09-902-540-7588	Sequence 7588, App
	37	46	2.5	7704	4	US-09-902-540-743	Sequence 743, App
C	38	45.8	2.5	34552	4	US-09-902-540-1262	Sequence 1262, App
	39	45.8	2.5	3556	4	US-09-902-540-9184	Sequence 9184, App
	40	45.6	2.5	534	4	US-09-902-540-982	Sequence 982, App
	41	45.6	2.5	576	4	US-09-252-991A-7886	Sequence 7886, App
	42	45.6	2.5	1047	4	US-09-252-991A-7886	Sequence 7886, App
	43	45.6	2.5	1566	4	US-09-902-540-4578	Sequence 4578, App
	44	45.6	2.5	1685	4	US-09-902-540-4578	Sequence 4578, App
	45	45.6	2.5	1696	4	US-09-252-991A-14548	Sequence 14548, App
	46	45.6	2.5	1695	4	US-09-252-991A-14548	Sequence 14548, App
	47	45.6	2.5	1686	4	US-09-252-991A-7816	Sequence 7816, App

ALIGNMENTS

```

RESULT 1
US-09-620-312D-494
; Sequence 494, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radojic T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 494
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1361)..(2419)
US-09-620-312D-494

```

	Query Match	22.9%	Score 421.4;	DB 4;	Length 3001;
	Best Local Similarity	63.8%	Pred. No. 8.4e-90;		
	Matches 677;	Conservative	0;	Mismatches 371;	Indels 13; Gaps 2;
QY	232	GGGAGGAGGCCGGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAAAATGCTGAGGCAA	291		
Db	1223	GGGATCGTGACAGGCCCTCCCTATGTGACGAGAGACAAGCAGGCCCAAGAAACAGGAGAAA	1282		

```
QY 292 AGCTCAG-GGGCCCTCCGGGGCAGCTTGTGGACATCGCGTGCMAAGTGTGCCAGGCGCTAC 350
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1283 AACCACAGTGTGTGTCTCCACAGAGTTTGTGGATGAAGCTCTGTGTGCGTCCGAGGAGTAC 1342
QY 351 CTGGGCGAGCTGAGCATGAGGACATCGACATCGTACGAGAGATGCGGTGAGGACCTCACT 410
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1343 CTTAGCAACTTGGCCACATGAGCATCGACAAGGACCTGGAGGCCCGGTGTACTCTCACC 1402
QY 411 GAGGCGGAGTGAGGAGACCTGACCCAGCAGTACTACTCTCCCTGGTTCATGCGCATGCTTTC 470
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1403 CCGGAGGGTGTGCTCTCTCCACAGCTACTACCAAGTGGTCCACGAAGGGGACAGA 1462
QY 471 ATCTCAATTAAGAAATTAATTTTCGAGTGCAGGCTCTGCTGAATAGAAATCAAGTCT 530
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1463 CTCAGGACCTCGACACTCAGGTTCAGCGCTGTGAGGACATCTGACGAGCTGCAGGCG 1522
QY 531 GTGAACCTTCAGACGACATTAAGCGGCTCCGGAACATCTGGATATTAAGGCCCGCGCC 590
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1523 GTGGTACCCACATAGACATGGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGAGCC 1582
QY 591 AAGTCCCGGGCGGAGACATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGCA 650
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1583 AAGTCCCGGGCGGAGGACATCATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGTG 1642
QY 651 GTGCGACACACCTCTTTCAGGGGACAAAGTGGTGGTCCAGAAGTACATCGAGAG 710
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1643 AACGGCAACCCCGTGTGTATGAAGGACGGCAAGTGGTGGTGCAGAAGTATATTGAGCGG 1702
QY 711 CGGCTGCTCATGTGACACCAAGTTCGACATCAGACAGTGGTCTCTGTCAGGACTGG 770
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1703 CCCCTCCTCATCTTTGGCACCAAGTTTGACCTCAGACAGTGGTCTCTGTTAACTGACTGG 1762
QY 771 AACCCCTCACCATCTGGTTTACAGGAGAGTTACTTGGGTTTCTCAACTCAGGCGCTTC 830
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1763 AACCACATTAACGTGTGGTTTACCGGACAGCTATATCCGGTTTTCACGCGAGCCCTTC 1822
QY 831 TCCTGGAACAAGTGACAGCGCATTCACCTGTGCAACAAAGCGGTCCAGAAATCACTCG 890
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1823 TCCTGGAAGAAGCTGACAACTCAGTGACCTGTGCAACAACTCCATCCAGAAGCACCTG 1882
QY 891 AAGATGATGTGGCGGACGCCCCCTGCTGCCGACACAAATGTGGACCAAGTGGTCTAGCCAGAGG 1942
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1883 GAGAACTCATGCGCATCGGCATCCACTGCTTCCGCCAGACAAATGTGGTCTAGCCAGAGG 1942
QY 951 TTCCAGGAGTACTCTGAGCGCCAGGGCCGCTGGCGCGTGTGGGCGAGCGTCACTACCCG 1010
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1943 TTCCAGGCCCACTGACGAGATGGTGGCCCCAATGCTTGTTCACCATCATCTGTGCT 2002
QY 1011 TCCATGAAGAAGGCCATCGCCCAACGATGAAGTGGCCCCAGGACCAAGTGAAGCTCGC 1070
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2003 GGCATGAAGGATGCTGTGATCCACGACCTTCAGACCTCCAGGACACCCGTGCAAGTTCGG 2062
QY 1071 AAGAACAGCTTTGAGCTCTACGGGCTGACTTCTGCTTGGAGGAGCTTCAGGCCCTGG 1130
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2063 AAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCTGTTTCGGGAGGAGCTTCAGGCCCTGG 2122
QY 1131 CTGATCGAGATCAATTCAGCCCCACCATGCAACCCGTCACCGCGGTCAACGCCCCAGCTG 1190
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2123 CTGATTGAGATCAACGCCAGCCCCAGATGGCAACCTTCACAGCAGTCACTGCCCCGGCTC 2182
QY 1191 TGTGACAGGTGAGGAGACACCAATCAAGGTGGCC-----GTGACCGCAGC 1238
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2183 TGTGCTGGGCTCAAGCTGACACCTTCGCCGTGGTCAATTGACCGGATGTGAGCGCCGAC 2242
QY 1239 TGTGACATCGGCACTTCAGACTCTGTGGAGGAGCGCGGT 1279
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2243 TGTGACAGAGGACCTTTGAGCTCATCTATAAGCAGCGCCGT 2283
```

RESULT 2

US-09-252-991A-5992

; Sequence 5992, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5992

; LENGTH: 909

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5992

Query Match

Best Local Similarity 3.0%; Score 56; DB 4; Length 909;

Matches 224; Conservative 44.4%; Pred. No. 0.0013;

Mismatches 280; Indels 0; Gaps 0;

QY 932 CATGTGGACGACAGGTTCCAGAGTACTTCAGCGCCAGGCGCGTGGCGCGTGG 991

Db 25 CCTGATGAGCTTCACCGGGCGCGCAACAGGAAGCGCGGTCCGGTGAAGTCCG 84

QY 992 GGGCAGCGTCACTACTCCGTCATGAAGAAGGCCATCGCCACGCGCATGAAGGTGGCCCA 1051

Db 85 CGTGGCGTGCAGGATATCTTACCGGGCTGTACTCTCCACCGCGGTGCTCGCGCCCT 144

QY 1052 GGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGTGTGACTTCGTCTTGG 1111

Db 145 GCGCCACCGCGACGTGAGCGGCATCGGCACGACATCGACATGGCCCTGCTCGAGGTGCA 204

QY 1112 GAGGACTTCAGGCCCTGGCTGATCGAGATCAATTTCCAGCCCCACCATGACCCGCTCCAC 1171

Db 205 GGTGCTGCTGCTGGCCAAACAGACCTTCAACTACTGTACCCGCGGTGCCCGCGGGG 264

QY 1172 GCGGCTCAGCGGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGGCGGTGGA 1231

Db 265 CTGGGCAACGACACACCCGGAACATCGTCCGTTACCGAGACTTCCCAACCGCCGACGGCA 324

QY 1232 CGCAGCTGTGACATTCGGCAACTTCGAGCTCTGTGAGGAGCGCGGTGTTGAGCGCC 1291

Db 325 CATGATCTCTACGTCGGCAACGACAGCAGTTCCGCAAGTTTCGCGAGCTGGCAGACCA 384

QY 1292 CCCATTCAGCGGTCGACCTCTGCGTGGCGGGCTGAGTGTGAGGAGAGCCAGGAGCA 1351

Db 385 CCGGAAATGGGCGGACGACCGCGCTTCGCCAACCAAGGCGCGGTGGCCAAACCGCA 444

QY 1352 GGTGCTGCGCGCTCTCAACCTCAAGGCTTCGGCTTCGCTTTGGACGGCGAGCGGTGAA 1411

Db 445 GGTGCTGATCCGCTGATCCGCGAGCCACGCTGCTGCACACCCGCGAGTGTATTCT 504

QY 1412 GGCACGGGCGCTTCGGCCATGCC 1435

Db 505 TTCCCTGGAGCGCGCGCGTGC 528

RESULT 3

US-09-252-991A-5951

; Sequence 5951, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5951
; LENGTH: 1305

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5951

Query Match 3.0%; Score 56; DB 4; Length 1305;

Best Local Similarity 44.4%; Pred. No. 0.0014;

Matches 224; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

QY 932 CATGTGACACAGCAGGTTCCAGGAGTACCTGACGCGCAGCGCGTGGCGCGTGTG 991

DB 546 CTTGATGAGCTCACCGGGCGCGGCAACAGAGAGGCGCGTCCGCTGAAGTGG 605

QY 992 GGGCAGCGTCTATCTACCCGCTCCATGAAGAGGCCATCGCCACGCCATGAAGTGGCCCA 1051

DB 606 CGTGGGCTGACCGATATCTCACCGGGCTGTACTCTCCACCGCGTGTCTCGCGCCCT 665

QY 1052 GAGCAGTGGAGCTCGAAGAACAGCTTTGAGCTCTAGGGGCTGACTTGTCTCTGG 1111

DB 666 CGCCCAACCGCGCTCAGCGGCGATCGGCCAGCACATCGACATGGCCCTGCTCGACGTGA 725

QY 1112 GAGGACTTCAGCGCTGCTGATCGAGATCAATTCAGCGCCCAACCATGCACCGTCCAC 1171

DB 726 GGTGCGCTGCTGGCCCAACAGACCCCTCAACTACCTGACCAACCGGCGTGGCGCGGCG 785

QY 1172 GCGGCTCACGGCCAGCTGTGTGCACAGGTGCAGGAGGACCATCAAGGTGGCGTGA 1231

DB 786 CTTGGGCAACGACACCCGAAACATCGTGCCTACAGGACTTCCCAACCGCGGCGGA 845

QY 1232 CCGCAGCTGTGACATCGGCAACTTGGAGTCTCTGTGGAGGACGCGGTGTGAGCGCG 1291

DB 846 CATGATCCTCAGCGTCGGCAACGACAGCCAGTTCGCAAGTTCGCGGAGCTGGCAGACCA 905

QY 1292 CCCATTACGCGGTCGACCTCTCGTGGCGGGCTCAGTGTGAGGAGCCAGGAGGCA 1351

DB 906 CCGGGAATGGCGACAGCCCGGCTTCGCCCAACAAAGCGCGGTGGCCCAACCGCGA 965

QY 1352 GGTGCTGCCGCTGTGCAACCTCAAGGCTCGGCGCTCGCTGTGGAGCGCGCGTGTGA 1411

DB 966 GGTGCTGATCCCGTGTATCCGCGAGGCCACGGTGTGCACACCGCCGCGGTGATCT 1025

QY 1412 GGCACGGGCGCTCTGGCCATGCC 1435

DB 1026 TTCCTGGAGCGCGCGCGTGC 1049

RESULT 4

US-09-252-991A-2471

; Sequence 2471, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2471

; LENGTH: 1272

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-2471

Query Match

Best Local Similarity 2.9%; Score 52.6; DB 4; Length 1272;

; Sequence 2632, Application US/09252991A

Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 545 GGCATTGACGGCTCCGGAACATCTGGATTATAAGCCCGCCAGTCCCGGGCCG 604

DB 756 GGAATGCTCGCGCAGCGCGGACGATTTCCGGCACCGCGAGGACGCGATCCT 815

QY 605 AGACATAGTGTGATCGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACC 664

DB 816 CGACGTACGATGGTTCGACCGCGGCGCGCTGGAGGTGTCGAGTAGGCCAGGC 875

QY 665 TCTTTCCAGGCAACAAGTGGTGTCCAGAAAGTACATCGAGACGCCGCTGTCTCATCTG 724

DB 876 GGTGACGCTGGAAGTTCGAGTTCGCCAGGACATCGAGCGGTGATCTCGGCTT 935

QY 725 TGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGGACCCCTGACCAT 784

DB 936 CATGATCAAGGACCGCTCGGCCAGCGGATGTACGGGATCAATACCCACCGCTGGCAA 995

QY 785 CTGGTTCTCAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 843

DB 996 GGGCTCACGACCTGAAGCGCGGAGCGCATCACTACCGCTTCGCTTCGACATGC 1054

RESULT 5

US-09-252-991A-2200/c

; Sequence 2200, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2200

; LENGTH: 1449

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-2200

Query Match 2.9%; Score 52.6; DB 4; Length 1449;

Best Local Similarity 48.5%; Pred. No. 0.0094;

Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 545 GGCATTGACGGCTCCGGAACATCTGGATTATAAGCCCGCCAGTCCCGGGCCG 604

DB 562 GGAATGCTCGCGCAGCGCGGACGATTTCCGGCACCGCGAGGACGCGATCCT 503

QY 605 AGACATAGTGTGATCGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACC 664

DB 502 CGACGTACGATGGTTCGACCGCGGCGCGCTGGAGGTGTCGAGTAGGCCAGGC 443

QY 665 TCTTTCCAGGCAACAAGTGGTGTCCAGAAAGTACATCGAGACGCCGCTGTCTCATCTG 724

DB 442 GGTGACGCTGGAAGTTCGAGTTCGAGGTTCGCCAGGACATCGAGCGGTGATCTCGGCTT 383

QY 725 TGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGGACCCCTGACCAT 784

DB 382 CATGATCAAGGACCGCTCGGCCAGCGGATGTACGGGATCAATACCCACCGCTGGCAA 323

QY 785 CTGGTTCTCAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 843

DB 322 GGGCTCACGACCTGAAGCGCGGAGCGCATCACTACCGCTTCGCTTCGACATGC 264

RESULT 6

US-09-252-991A-2632

; Sequence 2632, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2632
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2632

Query Match      2.9%; Score 52.6; DB 4; Length 2790;
Best Local Similarity 48.5%; Pred. No. 0.012;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 545 GCACATTGACGGCTCCGGAACATCTGGATTATAAGCCGCGGCAAGTCCCGGGCGG 604
Db 895 GGAATGCTCGCGGACCGCCAGGTACGGACGATTTCCGCGCACCGGCGAGCGCATCCT 954
QY 605 AGACATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCACCC 664
Db 955 CGACGTACGATGGTTCGACAGCGGCGCGGCTGGAGGTGTTCAGGTAGGCGAGC 1014
QY 665 TCTTTCCAGGGACAAACAGTGGTGTCCAGAGTACATCGAGACGCCCTGCTCATCTG 724
Db 1015 GGTGACGCTGGAAGTCGAGGTTCGAGGATTCGCCAGGACATCGAGCGCTCATCTCGGCTT 1074
QY 725 TGACACCAAGTTCGACATCAGACAGTGTCTCTCGTCAGGACTGGAACCCCTGACCAT 784
Db 1075 CATGATCAAGGACCGCTCGCGCAGCCGATGTACGGGATCAATACCCACCGCTTGACAA 1134
QY 785 CTGGTTCTACAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTCGACAAGC 843
Db 1135 GCGCTCACCAGCTGAAGCGCGGCGGAGCGCATCCTACCGCTTCGCTTCGACATGC 1193

RESULT 7
US-09-807-897-3
; Sequence 3, Application US/09807897
; Patent No. 6660851
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Yoshimitsu
; APPLICANT: UEKI, Jun
; TITLE OF INVENTION: NOVEL DNA FRAGMENT ENHANCING GENE EXPRESSION LEVEL
; FILE REFERENCE: 0230-0156P
; CURRENT APPLICATION NUMBER: US/09/807,897
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/JP00/05539
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Oryza sativa L.
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Variety: IR24 - Tissue: pistil
; OTHER INFORMATION: Library name: lambdaZAPII cDNA library derived from pistil mRNA
; OTHER INFORMATION: Clone name: RPC213
; NAME/KEY: misc feature
; OTHER INFORMATION: Strandedness: double - Topology: linear
; OTHER INFORMATION: Molecular type: cDNA to mRNA
; OTHER INFORMATION: Feature: nt1, nt2, nt3: transcription initiation points of RPC213
; OTHER INFORMATION: gene determined by primer extension
```

```
; NAME/KEY: misc feature
; OTHER INFORMATION: nt22-nt24: the first ATG of RPC213 gene
; OTHER INFORMATION: nt134-nt136: the second ATG of RPC213 gene
; OTHER INFORMATION: nt295-nt297: the third ATG of RPC213 gene
; OTHER INFORMATION: nt1276-nt1278: termination codon of RPC213 gene
; NAME/KEY: misc feature
; OTHER INFORMATION: nt1343-nt1348: nt1365-nt1370: poly (A) signals
; OTHER INFORMATION: nt1507-nt1524: poly (A)
US-09-807-897-3

Query Match      2.7%; Score 50.4; DB 4; Length 1524;
Best Local Similarity 47.2%; Pred. No. 0.032;
Matches 187; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 963 CTGCAGCGCCAGGGCGGTGGCGCGCTGTGGGGCAGCGTCATCTAGCCGCTCCATGAAGAAG 1022
Db 796 CAGCAAGCGGCGCGGCGCTGCCAAGATCTGCGGGGTGATCGCCGCGACGAGAGCGG 855
QY 1023 GCCATCGCCACGCCCATGAAGGTGGCCAGGACACAGTGGAGCCTCGCAAGAACAGCTTT 1082
Db 856 CACGAGGGCGGGCTACACGATGGCTCCGCGAGGCTGTTTCGAGCTCGACCCGAGCGCATG 915
QY 1083 GAGCTCTACGGGGCTGACTTCTCTTGGGAGGAGCTTCAGGCCCTTGGCTGATCGAGATC 1142
Db 916 GCGCGCGGCTGCGCGAGCTCATGCGGGGAAGGTGACCATCGCGGGGAGCTCATGTGCG 975
QY 1143 AATTCAGCCCCACCATGCACCCGCTCCAGCGCGTCAAGCGCCAGCTGTGTGCACAGGTG 1202
Db 976 GAGCGCGGCGGCGGCGGCGGAGCACAGCCTGTTC---GCCCGGTCTCCGCGGTGGCG 1032
QY 1203 CAGGAGGACACCATCAAGGTGGCCCTGGACCGCAGCTGTGACATCGGCAACTTTCGAGCTC 1262
Db 1033 GAGCGCGCGGCTGTACACGCGGAGGAGTACGCGGAACTCGTCGAGCACTTCTGTCGG 1092
QY 1263 CTGTGAGGCGACCGGTGTTGAGCGCGCCCTCATTCAGGGGTCCGACCTCTCGCTGGCG 1322
Db 1093 AGGTGCGGGGTGGCGGAGCTCGCGCGGGGCTCTCCGCGAGGCGCGACCGCGGAGGAG 1152
QY 1323 GCGGTGAGTGTCAGGAGAGCCAGGAGGAGGAGTGTG 1358
Db 1153 TACCTGTGGGGTGGCGCCCAAGATCCGAGGATG 1188

RESULT 8
US-09-902-540-983
; Sequence 983, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 983
; LENGTH: 9039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-983

Query Match      2.7%; Score 50.2; DB 4; Length 9039;
Best Local Similarity 46.9%; Pred. No. 0.065;
Matches 190; Conservative 0; Mismatches 213; Indels 2; Gaps 1;

QY 1241 TGACATCGCAACTTCGAGCTCTCTGTGAGGCGAGCGGTGGTTGAGCGCGCCCTTCAG 1300
Db 2988 TGACCTCGCTTCGATATATCCCGAGGAGGCTGCCAAGGAGCGTCCCGCGGTGAAGCA 3047
```

QY 1301 CGGGTCCGACCTCTGCGTGGCGGCGCTCAGTGTGAGGAGAGCCAGGAGCGAGTGTGCC 1360
Db 3048 AGCGTGTGGTCTGCGTGGCGGCGCTCTACTGACCGCTTGAAGGAAGAGCCCGTGGG 3107
QY 1361 CGTCTGCAACC--TCAAGGCGCTCGGCTCGCTGTGTGGACGCGCAGCGCTTGAAGGACCG 1418
Db 3108 AAGATGAGAGCCCATTCCTCCGCGCGCGGGGATGAAGGAGCTCCGCTCCCGCGGAAGCC 3167
QY 1419 GCGCCCTCGGCGCATGCCAGACCTTGCACGAGGACCCCATCACAGCTTTCAGCGGGAC 1478
Db 3168 GCTCCCGCGCTCGCGGAGGCCACCGCTGGAGCCCTCAAAGGTGGTCTCCCGTGAAG 3227
QY 1479 TTGGGACTGAAGGAAGAGAGGCGCTCCCTCTGCGCTTGTGACCCCTTAAAGGGGGCA 1538
Db 3228 ATTGCCAGGCGCGGCTCTGTATCGGCTTGACATCGAGGCGACCGCGCGCGGCC 3287
QY 1539 GCGGAGCGGCTGGAGCGGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 1598
Db 3288 GCGGAGCGGAGGAGGAGCG 3347
QY 1599 CGGCGCTGCGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1643
Db 3348 GCGCGCGTGGCG 3392

RESULT 9

US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-5

Query Match 2.6%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.097;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 833 CCTGGACAAGCTGGACAGCGCCATCCACCTGTGTGCAACAGCGCGTCCAGAAGTACCTGAA 892
Db 1437 CCAGAACATCTTCTGAGCGGCAACCGCGCTGTGCTGCAACATCGAGGCGCGCTGGA 1496

QY 893 GAATGATGTGGCGCGCAGCCCTGCTGCCGACACAAACATGTGGACCAACAGGTT 952
Db 1497 CGACAAGAGCTCGCAAGCTGCTGTGAGCGCCACAGCGGAGAGGCGTGCAT 1556
QY 953 CCAGGAGTACCTTCAGCGCGCGCGCTGTGGCGCGCTGTGGCGAGGCTCATCTACCGTC 1012
Db 1557 CAAGGAGTCCGCGTGTATGCGGACCTGAAGGCGTGCACGCAACATGAAGGCGCAGAG 1616
QY 1013 CATGAAGAGGCCATCGCCACCGCCATGAAGTGGCGCCAGGACCACTGTGAGGCTCGCAA 1072
Db 1617 CTTGGGCTAGCGCTTCGCGAGTTCCAGGAGCAGGACGCGCTGAAGGCGCTGCGCCT 1676
QY 1073 GAACAGC 1079
Db 1677 GATCAAC 1683

RESULT 10

US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-5

Query Match 2.6%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.097;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 833 CCTGGACAAGCTGGACAGCGCCATCCACCTGTGTGCAACAGCGCGTCCAGAAGTACCTGAA 892
Db 1437 CCAGAACATCTTCTGAGCGCGCGCGCTGTGCTGCAACATCGAGGCGCGCTGGA 1496

QY 893 GAATGATGTGGCGCGCAGCCCTGCTGCCGACACAAACATGTGGACCAACAGGTT 952
Db 1497 CGACAAGAGCTCGCAAGCTGCTGTGAGCGCCACAGCGGAGAGGCGTGCAT 1556

QY 953 CCAGGAGTACCTTCAGCGCGCGCGCTGTGGCGCGCTGTGGCGAGGCTCATCTACCGTC 1012
Db 1557 CAAGGAGTCCGCGTGTATGCGGACCTGAAGGCGTGCACGCAACATGAAGGCGCAGAG 1616

QY 1013 CATGAAGAAGCGCATCGCCACGCCCATGAAAGTGGCCCGCAGGACCAACGCTGGAGCCTCGCAA 1072
Db |||||
1617 CCTGGGCTACGCGCTTCGCGAGTTCCAGGAGCAGCAGCAGCGCCTGAAGCGCCTGCGCCT 1676
QY 1073 GAACAGC 1079
Db |||||
1677 GATCAAC 1683

RESULT 11
US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-098-487-5

Query Match 2.6%; Score 48.6; DB 2; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.097;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 833 CTGGACAAGCTGGACAGCGCCATCCCTGTGCAACACGCGCGTGCAGAACTACTGAA 892
Db |||||
1437 CCAGAACATCTTCGTGAGCGCAGCCGCGCTGTGCTGCACAACTGCCCCAAGCGCGTGA 1496
QY 893 GAATGATGTGGCGCGCAGCGCCCTGCTGCGCACACACATGTGGACCAACCAAGTT 952
Db |||||
1497 CGACAGCAGCTGCGCAAGCTGCTGTGAGCGCACACCGCGGAGAGGGCGTGCGCAT 1556
QY 953 CCAGAGTAGTACCTGCAGCGCCAGCGCGCGTGGCGCGCTGTGGGCGAGGCTCATCTACCCGTC 1012
Db |||||
1557 CAAGGAGTCCCGGTGATGCGCGACCTGAAGGCGGTGCACGCGACATGAAGGCCAGAG 1616
QY 1013 CATGAAGAAGGCCATCGCCACGCCCATGAAAGTGGCCCGCAGGACCAAGCTGGAGCCTCGCAA 1072
Db |||||
1617 CCTGGGCTACGCGCTTCGCGAGTTCCAGGAGCAGCAGCAGCGCCTGAAGCGCCTGCGCCT 1676
QY 1073 GAACAGC 1079
Db |||||
1677 GATCAAC 1683

RESULT 12
US-08-471-033-18
; Sequence 18, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Peetieidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note="Maize optimized DNA
; OTHER INFORMATION: sequence for VIPLA(a) 80 kd protein from AB78"
US-08-471-033-18

Query Match 2.6%; Score 48.2; DB 1; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.12;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 659 CCACCCCTCTTTCCAGGACAAAGTGGTGTCCAGAGTACATCGAGACGCCGCTGCT 718
Db |||||
720 CCACCCCTCTTCCAGGACAAAGTGGTGTCCAGAGTACATCGAGACGCCCATGAT 779

QY 719 CATCTGTGACACCAAGTTTCACATCAGACAGTGGTTCCTCGTCACGACTGGAAACCCCT 778
Db 780 GCTGGAGACCAACAGACCGCGCTCTACAGATCAAGACACCCACGCGACATCGT 839
QY 779 GACCATCTGTTTCAAGGAGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 838
Db 840 GACCGCGCGAGTGGAAACGCGTGATCCAGCAGATCAAGGCCAAGACCGCCAGCATCAT 899
QY 839 CAAGCTGGACAGCGCCATCCACCTGTGCAACAGCGCTCCAGAGTACCTGAGATGA 898
Db 900 CGTCGACGACGGGAGCGCTGGCCGAGAACGCGTGGCCGCGCCCAAGGACTACGAGAACCC 959
QY 899 TGTGGCGCGAGCGCCCTGCTGTCGCCGACACCAACATGTGACACCAAGTTCACAGGA 958
Db 960 CGAGGACAAACCCCACTGACCTTGAGGACCGCTGAGCTGAGCTACCCGACGA 1019
QY 959 GTACTCTGACGCGCAGCGCGCTGGCGCGTGTGGGCGAGGTCTATACCCGCTCCATGAA 1018
Db 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAAGAACAAAGCCCATCTACGAGAGCGGT 1079
QY 1019 GAAGGCATCGCCACCGCATGAGTGGTGGCCCGAGGACCAAGTGGAGCTCGCAAGACAG 1078
Db 1080 GATGACCTATCTAGACGAGAACCCGCAAGGAGGTGACCAAGCAGCTGAACGACACCC 1139
QY 1079 CTTTGAGCTTACGGGCTGACTTGTCTTGGAGGAGCTTACAGGCGCTGGCTGATCGA 1138
Db 1140 CGCAAGTTCAAGACGTGAGCCACCTGTACGAGGTGAAGCTGACCCCAAGATGAAGT 1199
QY 1139 GATCA 1143
Db 1200 GACCA 1204

RESULT 13

US-08-471-044-18
; Sequence 18, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA: US 08/037,057
; APPLICATION NUMBER: 25-MAR-1993
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SQLv3
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
; US-08-471-044-18
Query Match 2.6%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.12;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 659 CCACCCTCTTCCAGGAGCAACAAAGTGGTGGTCCAGAGTACATCGAGAGCGCGCTGCT 718
Db 720 CCACCCCATCACCTTGAACAAAGAGCGGTGGCAACCTCTGAACAAAGCCCATGAT 779
QY 719 CATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAAACCCCT 778
Db 780 GCTGGAGACCAACAGACCGCGCTCTACAAGATCAAGGACACCCACGCGACATCGT 839
QY 779 GACCATCTGTTCTCAAGGAGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 838
Db 840 GACCGCGCGAGTGGAAACGCGCTGATCCAGCAGATCAAGGCCAAGACCGCCAGCATCAT 899
QY 839 CAAGCTGGACAGCGCCATCCACCTGTGCAACAAAGCGGTCCAGAGTACCTGAGAGATGA 898
Db 900 CGTCGACGACGCGCGCGCTGGCCGAGAGCGCTGGCCGCGCCCAAGGACTACGAGAACCC 959
QY 899 TGTGGCGCGAGCGCCCTGCTGTCGCCGACACCAACATGTGACACCAAGCTTCCAGGA 958
Db 960 CGAGGACAGACCCCGAGCTGACCTTGAAGAGCGCTGAAGTGAAGCTACCCGACGA 1019
QY 959 GTACCTGACAGCGCAGCGCGCTGGCGCGTGTGGGCGAGCGTCTATCCCGCTCCATGAA 1018
Db 1020 GATCAAGGAGATCGAGGCGCTGCTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGT 1079
QY 1019 GAAGGCCATCGCCACCGCATGAGTGGCCCGAGGACCGAGCTCGCAAGAACAG 1078
Db 1080 GATGACCTATCTAGACGAGAACACCCGCAAGGAGGTGACCAAGCAGCTGAACGACACCC 1139
QY 1079 CTTTGAGCTTACGGGCGTGAATTCCTTCTTGGAGGAGCTTTCAGGCGCTGGCTGATCGA 1138
Db 1140 CGCAAGTTCAAGGAGCTGAGCCACCTGTACGAGGTGAAGCTGACCCCAAGATGAAGT 1199
QY 1139 GATCA 1143
Db 1200 GACCA 1204

RESULT 14

US-08-463-483A-18
; Sequence 18, Application US/08463483A
; Patent No. 5849870
; GENERAL INFORMATION:

840	840	GACCGCGCGGAGTGAAACGGCGTGATCCAGCAGATCAAGGCCAAGACCCGAGCATCAT	899
839	839	CAAGCTGGACAGCGCCATCCACCTGTGTCAACAAACGCGTCCAGAAAGTACTGTGAAGAATGA	898
900	900	CGTGCAGCAGCGGCGCGTGCGCCGAGAAGCGGTGGCGCGCAAGGACTACGAGAAACCC	959
899	899	TGTGGGCGGACAGCCCTGCTGCTGCCCGGACACAAACATGTGGACACAGCACCAGGTTCCAGGA	958
960	960	CGAGGACAAGACCCCAAGCCTGACCCCTGAAGGACGCGCTGAAGCTGAGCTTACCCCGACGA	1019
959	959	GTAACCTGCAGCGCCAGGCGCGTGTCGGGCGAGCGTCTTACCCGTCCTATG	1018
1020	1020	GATCAAGAGAGATCGAGGGCTTGTGTACTACAAGAACAGGCCCATCTACAAGAGCAGCGT	1079
1019	1019	GAAGGCGCATCGCCACGACCCATGAAGGTGGCCCAAGGACACAGTGGAGCCTCGCAAGAACAG	1078
1080	1080	GATGACCTATCTAGACGAGAACACCCCAAGGAGGTGACCAACGAGCTGAACGACACCCAC	1139
1079	1079	CTTTGAGCTCTACGGGGTGACTTCGTCTCTTGGGAGGAGCTTCAAGGCCCTGGCTGATCGA	1138
1140	1140	CGCAAGATTCAAGGAGCTGAGCCACCTGTACGACGTTGAAGCTGACCCCAAGATGAACGT	1199
1139	1139	GATCA 1143	
1200	1200	GACCA 1204	
RESULT 15			
US-08-471-046A-18			
; Sequence 18, Application US/08471046A			
; Patent No. 5866326			
; GENERAL INFORMATION:			
; APPLICANT: Warren, Gregory W			
; APPLICANT: Koziel, Michael G			
; APPLICANT: Mullins, Martha A			
; APPLICANT: Nye, Gordon J			
; APPLICANT: Carr, Brian			
; APPLICANT: Desai, Nalini M			
; APPLICANT: Kostichka, N. Kristy			
; APPLICANT: Duck, Nicholas B			
; APPLICANT: Estruch, Juan J			
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal			
; TITLE OF INVENTION: Protein Genes			
; NUMBER OF SEQUENCES: 50			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: No. 5866326artis Corporation			
; STREET: 3054 Cornwallis Road			
; CITY: Research Triangle Park			
; STATE: NC			
; COUNTRY: USA			
; ZIP: 27709			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30B			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/471,046A			
; FILING DATE: 06-JUN-1995			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/463,483			
; FILING DATE: 05-JUN-1995			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/314,594			
; FILING DATE: 09-SEP-1994			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/218,018			
; FILING DATE: 23-MAR-1994			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/037,057			
; FILING DATE: 25-MAR-1993			
; ATTORNEY/AGENT INFORMATION:			

	Query Match Best Local Similarity Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;	2.6%; Score 48.2; DB 2; Length 2004; Pred. No. 0.12;
QY	659 CCACCCCTCTTTCAGGACAAAGTGGTGGTCCAGAAGTACATCGAGACGCCGCTGCT 718 	
Db	720 CCACCCCATCACCTGAACAAGAAGCAGTGGGACAACCTGCTGAACAACAAGCCCATGAT 779 	
QY	719 CATCTGTGACACAAAGTTGCAATTCAGACAGTGGTTCTCGTCAGGACGTGGAAACCCCT 778 	
Db	780 GCTGGAGACCAACCAGACCGACGGCGTCTACAAGATCAAGGACACCCACGCAACATCGT 839 	
QY	779 GACCATCTGTTCTTCAAGAGAGATTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTCGGA 838 	

```
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CCI695/CIP3/DIV8 - SQLv4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1(a) 80 kd protein from AB78"
; US-08-471-046A-18

Query Match      2.6%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.12;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 659 CCACCCCTCTTCCAGGAGACAACAGTGGGTGCTCCAGAAATACATCGAGACGCCGCTGCT 718
Db      |||||
QY 720 CCACCCCATCACCTGAAAGAACAGAGGTGGACACCTGCTGAACAAAGCCCATGAT 779
Db      |||||
QY 719 CATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCACGAGCTGGAAACCCCT 778
Db      |||||
QY 780 GCTGGAGACCAACAGACCCGCGCTTCAAGATCAAGGACACCCACCGCAACATCGT 839
Db      |||||
QY 779 GACCATCTGCTTCTACAAGGAGAGTTACTTTCGGTCTCTCAACTCAGCGCTTCTCCCTGGA 838
Db      |||||
QY 840 GACCGCGCGAGTGGAAACGGCTGATCCAGCAGATCAAGGCCAAGACCCGACGATCAT 899
Db      |||||
QY 839 CAAGCTGGACAGCGCCCATCCACCTGTGTCACACAGCGCGTCCAGAAGTACCTGAAGATGA 898
Db      |||||
QY 900 CGTCGACGACGGCGAGCGCGTGGCCGAGAGAGCGGTGGCCGCAAGGACTACGAGAACCC 959
Db      |||||
QY 899 TGTGGCGCGCAGCCCTGCTGTCGCCGACACACATGTGGACCAACAGGTTCCAGGA 958
Db      |||||
QY 960 CGAGGACAAGACCCCGAGCTGACCTTGAAGGACGCGCTTGAAGCTGAGCTACCCCGACGA 1019
Db      |||||
QY 959 GTACCTGCAGCGCAGGCGCGTGGCGCGTGTGGGGCAGCGCTCATCTACCCGTCCTGAA 1018
Db      |||||
QY 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAGAACAGCCCATCTACGAGAGCAGCGT 1079
Db      |||||
QY 1019 GAAGGCGCATCGCCACGCCATGAAGGTGGGCCCGAGGACCAAGTGGAGCTCGCAAGAACAG 1078
Db      |||||
QY 1080 GATGACCTATCTAGACCGAGAACACCGCAAGGAGGTGACCAAGCAGCTGAACGACACCCAC 1139
Db      |||||
QY 1079 CTTTGAGCTTACGGGCTGACTTCGTCCTTGGGAGGGAATTGAGGCCCTGGCTGATCGA 1138
Db      |||||
QY 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
Db      |||||
QY 1139 GATCA 1143
Db      |||||
QY 1200 GACCA 1204
Db      |||||
```

Search completed: September 24, 2005, 11:30:34
Job time : 337.858 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:08:28 ; Search time 7063.48 Seconds
(without alignments)
9904.766 Million cell updates/sec

Title: US-10-615-659-1
Perfect score: 1838
Sequence: 1 tgggagcagctggggcccc.....actctcccaaggcggaattc 1838

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1196	65.1	1259	9 AY415398	Homo sapi
2	1145.6	62.3	1257	9 AY415399	Pan trogl
3	818.8	44.5	2969	3 AK030151	Mus muscu
4	812.6	44.2	2720	3 AK029745	Mus muscu
5	768	41.8	770	7 CR745100	CR745100
6	767	41.7	1265	9 AY415400	Mus muscu
7	443.6	24.1	775	5 BU937372	AGENCOURT
8	421.4	22.9	4184	3 HSM808322	AGENCOURT
9	412.4	22.4	777	7 CK597795	Homo sapi
10	399	21.7	2520	3 AK080321	Mus muscu
11	378.4	20.6	1023	4 BM808637	AGENCOURT
12	360	19.6	1022	4 BM808516	AGENCOURT
13	321	17.5	693	4 BM717061	UI-E-EJO-
14	316	17.2	861	7 CK653583	AGENCOURT
15	307.6	16.7	896	6 CD515431	AGENCOURT
16	298.4	16.2	762	6 CD627148	AGENCOURT
17	284.2	15.5	503	7 CK604006	AGENCOURT
18	258.8	14.1	583	7 CN349423	170004243
19	258.2	14.0	680	6 CB854808	UI-CF-EC1
20	252.8	13.8	514	7 CR746623	CR746623
21	245.4	13.4	320	1 AA144756	mr73404.r
22	231.4	12.6	406	1 AI026623	ov97b08.x
23	222.4	12.1	429	7 W31827	zb94f04.r1
24	219.2	11.9	880	5 BX448465	BX448465

25	213.2	11.6	616	7	CN349424	CN349424
26	205	11.2	453	7	W31397	W31397
27	200.8	10.9	963	5	BX431541	BX431541
28	200.6	10.9	580	5	BP305902	BP305902
29	199	10.8	590	1	AL045121	DKF2p434M
30	195.6	10.6	484	2	BF830679	RC3-HT023
31	194.2	10.6	3328	3	BC028169	BC028169
32	192.2	10.5	414	7	CN096245	EC2CA11A
33	191.6	10.4	724	4	BG916644	BG916644
34	191.4	10.4	875	5	BQ718874	AGENCOURT
35	190.4	10.4	651	7	CF999072	CF999072
36	188	10.2	504	5	BQ327883	MR4-RT004
37	183.6	10.0	608	7	CN349425	CN349425
38	176.6	9.6	498	1	AA551103	AA551103
39	173.6	9.4	488	9	CE760083	tigr-gss8
40	172.4	9.4	482	2	BE102364	UI-R-BT1-
41	167.6	9.1	812	6	CD648578	AUF_103_C
42	161.8	8.8	977	5	BX431542	BX431542
43	158.8	8.6	630	2	BE865183	BE865183
44	157.2	8.6	697	7	CK961833	CK961833
45	156	8.5	850	5	BX849970	BX849970

ALIGNMENTS

RESULT 1
AY415398
LOCUS
DEFINITION Homo sapiens HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
AY415398 1259 bp DNA linear GSS 12-DEC-2003
ACCESSION AY415398
VERSION AY415398.1 GI:39771357
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1259)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1259)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1259
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1259
/locus_tag="HCM5544"

ORIGIN
Query Match 65.1%; Score 1196; DB 9; Length 1259;
Best Local Similarity 100.0%; Pred. No. 2.2e-252;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 136 AAGACTTCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCT 195
DB 1 AAGACTTCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCT 60

```
QY 196 GCAGCAGGAGCAGCAAGAACAGCCAGGGAACAGAGGAGGAGGCCCGGAGCAGCGACC 255
Db |||||||
61 GCAGCAGGAGCAGCAAGAACAGCCAGGGAACAGAGGAGGAGGCCCGGAGCAGCGACC 120
QY 256 TGAGCAGCAGGCAAGATGCTGAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGC 315
Db |||||||
121 TGAGCAGCAGGCAAGATGCTGAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGC 180
QY 316 TTGTGACATCGCGTCAAGGTGTGTCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACA 375
Db |||||||
181 TTGTGACATCGCGTCAAGGTGTGTCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACA 240
QY 376 TCGACACGTGAGGAGGAGCTTCACTGAGGCGGAGTGGAGGAGCTTGACCC 435
Db |||||||
241 TCGACACGTGAGGAGGAGCTTCACTGAGGCGGAGTGGAGGAGCTTGACCC 300
QY 436 AGCAGTACTACTCCCTCGTTTCATGGGATGCTTTCATCTCCAATTCAGAAATTAATCTTT 495
Db |||||||
301 AGCAGTACTACTCCCTCGTTTCATGGGATGCTTTCATCTCCAATTCAGAAATTAATCTTT 360
QY 496 CGCAGTGCCAGGCTCTGCTGAATAGAAATCAGCTCTGTGAACCTCAGACGGACATTTGAGC 555
Db |||||||
361 CGCAGTGCCAGGCTCTGCTGAATAGAAATCAGCTCTGTGAACCTCAGACGGACATTTGAGC 420
QY 556 GGCTCCGGAACATCTGGATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATATGTT 615
Db |||||||
421 GGCTCCGGAACATCTGGATTATAAGCCCGCGCCAAAGTCCCGGGCCGAGACATATGTT 480
QY 616 GCATGACCGTGTGAGGAGATCTCGAGCTGGCAGCTGCACACACCTCTTTTCAGGG 675
Db |||||||
481 GCATGACCGTGTGAGGAGATCTCGAGCTGGCAGCTGCACACACCTCTTTTCAGGG 540
QY 676 ACAACAAGTGGTGTTCAGAAAGTACATCAGAGCGCGCTGCTCATCTGTGACACCAAGT 735
Db |||||||
541 ACAACAAGTGGTGTTCAGAAAGTACATCAGAGCGCGCTGCTCATCTGTGACACCAAGT 600
QY 736 TCGACATCAGACAGTGGTTTCTCGTCACGAGCTGGAACCCCTGACCATCTGTTCTACA 795
Db |||||||
601 TCGACATCAGACAGTGGTTTCTCGTCACGAGCTGGAACCCCTGACCATCTGTTCTACA 660
QY 796 AGGAGATTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGGCCA 855
Db |||||||
661 AGGAGATTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGGCCA 720
QY 856 TCCACTGTGCAACACCGCTCCAGAGTACTGAAAGTATGATGTGGCGCGCAGCCCC 915
Db |||||||
721 TCCACTGTGCAACACCGCTCCAGAGTACTGAAAGTATGATGTGGCGCGCAGCCCC 780
QY 916 TGCTGCCCGCACACAATGTGGACAGCACAGGTTCCAGGAGTACCTGACGCGCCAGG 975
Db |||||||
781 TGCTGCCCGCACACAATGTGGACAGCACAGGTTCCAGGAGTACCTGACGCGCCAGG 840
QY 976 GCGCTGGCGCGCTGTGGGCGAGGTCTATCCGCTCATGAAGAGGCGCATCGGCCAGG 1035
Db |||||||
841 GCGCTGGCGCGCTGTGGGCGAGGTCTATCCCGCTCATGAAGAGGCGCATCGGCCAGG 900
QY 1036 CCATGAAGTGGCCAGGACACGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGG 1095
Db |||||||
901 CCATGAAGTGGCCAGGACACGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGG 960
QY 1096 CTGACTTCGTTGGAGGGGACTTCAGGCCCTTGGCTGATCGAGATCAATTCAGGCCCA 1155
Db |||||||
961 CTGACTTCGTTGGAGGGGACTTCAGGCCCTTGGCTGATCGAGATCAATTCAGGCCCA 1020
QY 1156 CATGACCCCGTCCAGCCCGGTCAAGGCCAGCTGTGTGACAGGTGAGGAGGACACA 1215
Db |||||||
1021 CATGACCCCGTCCAGCCCGGTCAAGGCCAGCTGTGTGACAGGTGAGGAGGACACA 1080
QY 1216 TCAAGTGGCGCTGGACCGCAGCTGTGACATCGGAACTTCAGAGCTCCTGTGAGGAGCAGC 1275
Db |||||||
1081 TCAAGTGGCGCTGGACCGCAGCTGTGACATCGGAACTTCAGAGCTCCTGTGAGGAGCAGC 1140
```

```
QY 1276 CGGTGGTTGAGCGCGCCCAATTCAGCGGTCGACACTCTGCTGGCGGCGTCAGT 1331
Db |||||||
1141 CGGTGGTTGAGCGCGCCCAATTCAGCGGTCGACACTCTGCTGGCGGCGTCAGT 1196

RESULT 2
AY415399 1257 bp DNA linear GSS 12-DEC-2003
LOCUS Pan troglodytes HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY415399
VERSION AY415399.1 GI:39771358
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 1257)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1257)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1257
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1257
/locus_tag="HCM5544"
gene
ORIGIN
Query Match 62.3%; Score 1145.6; DB 9; Length 1257;
Best Local Similarity 96.2%; Pred. No. 2.7e-241;
Matches 1151; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 136 AAGACTTCGGCGCACCATGGCATCCAGTGGGTGGTTCAGCCACCGAGCT 195
Db |||||||
1 AAGACTTCGGCGCACCATGGCATCCAGTGGGTGGTTCAGCCACCGAGCT 60
QY 196 GCAGCAGGAGCAGAGAACAGCCAGGACAGAGGAGGAGGCCCGGAGCAGCGACC 255
Db |||||||
61 GCAGCAGGAGCAGAGAACAGCCAGGACAGAGGAGGAGGCCCGGAGCAGCGACC 120
QY 256 TGAGCAGCAGCAAGATGCTGAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGC 315
Db |||||||
121 TGAGCAGCAGCAAGATGCTGAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGC 180
QY 316 TTGTGGACATCGCGTGAAGGTGTGCCAGGCTACCTGGGGCAGCTGGAGCATGAGGACA 375
Db |||||||
181 TTGTGGACATCGCGTGAAGGTGTGCCAGGCTACCTGGGGCAGCTGGAGCATGAGGACA 240
QY 376 TCGACACGTGAGGAGGAGCTTCACTGAGGCGGAGTGGAGGAGCTTGACCC 435
Db |||||||
241 TCGACACGTGAGGAGGAGCTTCACTGAGGCGGAGTGGAGGAGCTTGACCC 300
QY 436 AGCAGTACTACTCCCTCGTTTCATGGGATGCTTTCATCTCCAATTCAGAAATTAATCTTT 495
Db |||||||
301 AGCAGTACTACTCCCTCGTTTCATGGGATGCTTTCATCTCCAATTCAGAAATTAATCTTT 360
QY 436 AGCAGTACTACTCCCTCGTTTCATGGGATGCTTTCATCTCCAATTCAGAAATTAATCTTT 495
Db |||||||
301 AGCAGTACTACTCCCTCGTTTCATGGGATGCTTTCATCTCCAATTCAGAAATTAATCTTT 360
```

Qy	496	CGCAGTGCACGGCTCTGCTGAATAGAATCACGTCGTGTAACCCCTCAGACGGACATTCAGC	555
Db	361	CGNANNNCCAGGCTCTGCTGAANNNNNTTCANNNTNTNNNCNTNAGACGGACATTCAGC	420
Qy	556	GGCTCCGGACATCTGGATTATTAAGCCCGCGCCCAAGTCGCCGGGCGGACATAGTGT	615
Db	421	GGNNCCGGAAATCTCGATTATAAAGCCCGCGGCAAGTCGCCGGGCGGACATAGTGT	480
Qy	616	GCATGGACCGGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCTCTTTCACGGG	675
Db	481	GCATGGACCGGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCTCTTTCACGGG	540
Qy	676	ACAACAAGTGGGTGGTCAGAAATACATPCGAGACGCCGCTGCTCATCTGTGACACCAAGT	735
Db	541	ACAACAAGTGGGTGGTCAGAAATACATPCGAGACGCCGCTGCTCATCTGTGACACCAAGT	600
Qy	736	TCGACATCAGACAGTGGTTCTCGTCACGGAAGTGGAAACCCCTGACCATCTGGTCTTACA	795
Db	601	TCGACATCAGACAGTGGTTCTCGTCACGGAAGTGGAAACCCCTGACCATCTGGTCTTACA	660
Qy	796	AGGAGAGTTACTTTCGGGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCA	855
Db	661	AGGAGAGTTACTTTCGGGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCA	720
Qy	856	TCCACCTGTGCAACAAACCGCGTCCAGAAATGATCTGGAAGTGTGGGCGCGACGCCCC	915
Db	721	TCCACCTGTGCAACAAACCGCGTCCAGAAATGATCTGGAAGTGTGGGCGCGACGCCCC	780
Qy	916	TGTCGCCCGCACACAAATGTGGACACGACCCAGGTTCCAGGAGTACTTGCAGGCGCAGG	975
Db	781	TGTCGCCCGCACACAAATGTGGACACGACCCAGGTTCCAGGAGTACTTGCAGGCGCAGG	840
Qy	976	GCCGTGGCGCGTGTGGGGCAGCGTCACTACCCGTCATGAAAGAGCGCATCGCCACCG	1035
Db	841	GCCGGGCGCGTGTGGGGCAGCGTCACTACCCGTCATGAAAGAGCGCATCGCCACCG	900
Qy	1036	CCATGAAGTGGGCCAGGACCAAGTGGAGCGCTTCGCAAGAACAGCTTTGAGCTCTACGGG	1095
Db	901	CCATGAAGTGGGCCAGGACCAAGTGGAGCGCTTCGCAAGAACAGCTTTGAGCTCTACGGG	960
Qy	1096	CTGACTTCGTCTTGGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCA	1155
Db	961	CTGACTTCGTCTTGGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCA	1020
Qy	1156	CCATGCAACCCGTCACGCGCGTCCAGCCAGCTGTGTGCACAGTGCAGGAGGACACCA	1215
Db	1021	CCATGCAACCCGTCACGCGCGTCCAGCCAGCTGTGTGCACAGTGCAGGAGGACACCA	1080
Qy	1216	TCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCCTGTGGAGGAGC	1275
Db	1081	TCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCCTGTGGAGGAGC	1140
Qy	1276	CGTGGTTGAGCGCGCCCAATTCAGCGGTTCGACCTCTGGTGGCGGGCGTCACT	1331
Db	1141	NGTGGTTGAGCGCGCCCAATTCAGCGGTTCGACCTCTGGTGGCGGGCGTCACT	1196
RESULT 3			
AK030151		2969 bp	linear
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			

```
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
490..2835
/notes="unnamed protein product; hypothetical
Tubulin-tyrosine ligase containing protein
(InterPro:IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26811.1"
/db_xref="GI:26326135"
/tranlation="MENGRRKLSSTLSGDGHHKKNLKGIPQDLSLSPKLDKRIA
RQTEKAIKERKFIYIGHYPIVATLRKRWVEKFNFFPKALQNLGSEDSKAEKE
NGEIALRDEDDTHDMSRVXNEIPYLLWITIKRDVVDYHSLDCDQMLNHYGKTAFTT
KIGLCINLMRSLFWYQANPTFFPCYGLCTSEKEQDFLDFFRTVAASILKWVVLHQ
NYCSKVGSKKKEAKNSDPSKQDPENPDLKPLSLGOVVDCTACKVCAOYLQLEHE
DIDVSEASTEALSEERNWDLTOQYLLVHGNASITDSKSYFAQCQALLSLSSVNPOT
EIDGIRNIWIIKPAKSRGRDIVCMRDVENILSLVAUSDQTTKDNKVVQKIEPML
IYDTFDIRKQWFLVTDNPLTIWFKESYLRSTQRFSLDKLOSALHLCNNSIQRRLK
NDKERSPLLPCHNMWTSFQYLOKRGGRGTWSIIYPSMKRAVTNMRVAQDREA
RKNSFELYGADFILGRDFPWLIEINSSPTHPSTPTAQLCAQVQEDTIIKVVDRKL
DRNCIDIGPELLRQPAVELPFPNGSDLCVEGISVKKAKKOMPPIASVGLSLSLDAP
PKVRSARALMETVIRPPTTVRODMKREAKVLSTTWSMPVMDAENVRGRAKPIYAFV
NDYQVNDKNSKSGYTRVSSKVPYGVTLTSAQHPALPAQTMKPTQMTSPPTASGNH
RUSDFPCPIVFBEELMLHPSNRPSRCILQSRAGWIRGIP"
```

ORIGIN

```
Query Match      44.5%; Score 818.8; DB 3; Length 2969;
Best Local Similarity 74.8%; Pred. No. 2.5e+169;
Matches 1075; Conservative 0; Mismatches 342; Indels 21; Gaps 3;

QY 130 TTTCAAGACGCTCCGGCGCACCATGGCATCCAGCATCCTCAAGTGGGTGGTCAGCCACC 189
DB 1082 TCTTGGATGACTTCGGCGCACGGTGGCTGCCAGCATCCTCAAGTGGGTAGTCTTACC 1141

QY 190 AGAGCTGCAGCAGGACGACGAGNAGCAAGCCAGGACCCAGAGGAGGAGCGGGGAGCA 249
DB 1142 AGAATTACT-----GCAGCAAGTCAAGGCAAGAGTAGAAGAGGAGGAGGCAAGCA 1195

QY 250 GCGACCTGAGCAGCAGCGCAAGATGCTGAAAATGCTGAGGCAAGCTCAGGGGCTCCCGG 309
DB 1196 GTGACCCAGCCAGCCAGAAAGATCTGAAAATCCTGACCTCAAGTACCAAGCTCTCG 1255

QY 310 GCGAGCTTGTGACATCGCGTGCAAGGTGTGCGAGGCTTACCTGGGCGCAGCTGGAGCATG 369
DB 1256 GACAGGTTGTAGACACAGCTTGCAAGGTGTGCGAGGCTTACCTGGGACAGCTAGAGCATG 1315

QY 370 AGGACATCAGACGCTCAGCAGATCGCGTGAGGACCTCACTGAGCGCGAGTGGGAGGACC 429
DB 1316 AAAGACATAGACGTATCGGAGGCGCAGCACCGAGGCTTGTCTGAGGAGGAATGGAATGACC 1375

QY 430 TGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAATT 489
DB 1376 TGACACAGCAGTACTACTTCTGTTTCATGGCAATGCTTCCATCACTGATTCGAAAAGTT 1435

QY 490 ACTTTTGCAGTGCCAGGCTCTCTGTAATAGAAATCAGTCTGTGAACCCCTCAGACGAGCA 549
DB 1436 ACTTTGCGCAATGCCAAGCTCTGCTGAGTAAGATCAGTTCTGTGAACCCCTCAGACAGAGA 1495

QY 550 TTGACGGGCTCGGAACATCTGATATTAAGCCCGCGGCGCAAGTCCCGGGCGGAGACA 609
DB 1496 TCGATGGGATCGGGAACATCTGATATTAAGCCTTGACGCAAGTCCCGGGGTTCGAGATA 1555

QY 610 TAGTGTGATGACCGCTGTGGAGGAGATCTCTGAGCTGCAGCTGCAGACCACTCTTT 669
DB 1556 TTGTGTGTATGACCGCTGTGGAGAACATCTGAGTCTGTGTGGCGGAGACAGCAGACCA 1615

QY 670 CCAGGACAAACAGTGGGTGGTCCAGAAGTACATCGAGACGCGGCTGTCTCATCTGTGACA 729
DB 1616 CAAAGGACAAACAAATGGGTGGTCCAGAAGTACATCGAGACCCCAATGTCTCATCTAGACA 1675

QY 730 CCAAGTTCACATCAGACAGTGGTCTCTGTACGAGCTGGAACCCCTGACCATCTGGT 789
```

```
DB 1676 CCAAGTTTGATCATGACAGACAGTGGTTCCTCGTCAAGATCGGAATCCCCTAACCATCTGGT 1735
QY 790 TCTACAAGGAGAGTGTACTTTCGGGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGAGCTGGACA 849
DB 1736 TCTACAAAGAAAGTACTACCTCCGCTTCTCCACACAGGGCTTCTCCCTGGACAAGTGGACA 1795
QY 850 GGGCCATCCACTGTGTGCAACAAACGGCGTCCAGAAGTACTCTGAAGAATGATGTGGGCGGCA 909
DB 1796 GGGCCATCCACTGTGTAAACAACTCATCCAGAGGCGCTCAAAATGATAAAGAGCGCA 1855
QY 910 GCCCCTCTGTCGCCCCACACACATGTTGGACACAGACACAGGTTCCAGGAGTACTCTGCAGC 969
DB 1856 GTCCGCTGTCTACTTGTGTATAATGTTGGACACAGACTCGTTTTCAGGAGTACCTTCGACA 1915
QY 970 GCCAGGGCGGTGGCGCGCTGTGGGCGAGCGTCTATCTACCCGTCCTATGAAGAAGAGCCATCG 1029
DB 1916 AGAGGGGCGGAGGAGAAAGTGGGGTAGCATCATCTACCCGCTATGAAAAGAGCTGTCA 1975
QY 1030 CCCACGCCATGAAGGTGGGCCACAGGACCGAGTGGAGCTCGCAAGAACAGCTTTGAGCTCT 1089
DB 1976 CCAATGCCATGAGGGTAGCCACAGGACACGTAGAAGCCCGTAAGAAACAGCTTCGAGCTCT 2035
QY 1090 ACGGGGCTGACTTCGTCCTTGGGAGGAGCTTTCAGGCGCTGGCTGATCCGAGATCAATTCCA 1149
DB 2036 ACGGAGCGGACTTCATCTTGGGGCGAGACTTCAAGCATGGCTTATCGAGATCAACTCCA 2095
QY 1150 GCCCCACCATGACCCCGTCCACGCGGTCCAGCGGCGAGCTGTGTGCACAGGTGCAGGAGG 1209
DB 2096 GCCCCACCATGACCCCGTCCACTCCGTCACAGCTCAGCTCTGTGCCAGGTGCAGGAGG 2155
QY 1210 ACACATCAAGGTGGCGGTGGACCGC-----AGCTGTGACATCGGCAACTTCG 1257
DB 2156 ACACATCAAGGTGGTGGATCGATCGAAACTGGAACGAACTGTGACATCGGCAACTTTG 2215
QY 1258 AGCTCCTGTGAGGAGCAGCGGTGTGAGCGCGCCCAATTCAGCGGTCGACCTCTCG 1317
DB 2216 AGCTTCTGTGGGCGGAGCGCTGCCGAGCTGCCACCGTTCAACGGGTCTGACCTCTGTG 2275
QY 1318 TGGCGGGCGGTCAGTGTGAGGAGGAGCAGGAGCAGGTGTGCGCCGCTGSCAACCTCAAGG 1377
DB 2276 TGGAGGTATCAGCGTGAAGAAAGCCAGAACTCCGCGCCATTCGACGCTCGGCC 2335
QY 1378 CTTGGCGCTTGTGTGAGCGGAGCGCTGAAGCGACGGGGCCCTCTCGGCATCGCCAG 1437
DB 2336 TCTCAGAATCACTCTTAGATG---TCCACCCAAATGCGGAGTGCCTGGGCGCTGATGG 2392
QY 1438 ACCCTGCCAGGACCCCATCACAGCTCTCCACGGGACTTGGGACTTGAAGGAAGAGA 1497
DB 2393 AAACAGTATCAGGCTCCAGAACGACGTCGGCAGACTTGGAAACGGGAAGAGCAA 2452
QY 1498 AGGGGCTCCCCCTGGCCTTGTGTGGCACCCCTTAAGGGGGGCGACCGGAGCGGTGGAGC 1555
DB 2453 AGGTACTTTCACCACTTGGTCCATGCCAGTAAATGATGCAGAGGTTCAGAGGTTCGAGC 2510

RESULT 4
AK029745
LOCUS
DEFINITION
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4930524K07 product: hypothetical Tubulin-tyrosine
ligase containing protein, full insert sequence.
AK029745
AK029745.1 GI:26325681
HTC; CAP trapper.
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
Carninci, P. and Hayashizaki, Y.
AUTHORS
High-efficiency full-length cDNA cloning
TITLE
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
99279253
MEDLINE
```

PUBMED REFERENCE AUTHORS	10349636 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBMED	Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
PUBMED REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuoka, M., Yoneda, Y., Ishikawa, T., Osawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer
JOURNAL MEDLINE PUBMED	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
PUBMED REFERENCE AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL REFERENCE	Nature 409, 685-690 (2001) 5
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL REFERENCE AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs 6 (bases 1 to 2720) Nature 420, 563-573 (2002)
TITLE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/
FEATURES source	Location/Qualifiers 1. .2720 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:4930524K07" /db_xref="taxon:i0090" /clone="4930524K07" /sex="male" /tissue type="testis" /clone_id="RIKEN full-length enriched mouse cDNA library"


```

Qy 943 GCACAGGTTCCAGAGTACCTGACGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGTCA 1002
Db |||||
541 GCACAGGTTCCAGAGTACCTGACGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGTCA 600
Qy 1003 TCTACCGTCCATGAAGAAGCCATCGCCACGCGCATGAAGTGGCCAGGACCAACGCTGG 1062
Db |||||
601 TCTACCGTCCATGAAGAAGCCATCGCCACGCGCATGAAGTGGCCAGGACCAACGCTGG 660
Qy 1063 AGCTTCGCAAGAACAGCTTTGAGTCTACCGGGGTGACTTCGTCTTTGGGAGGACTTCA 1122
Db |||||
661 AGCTTCGCAAGAACAGCTTTGAGTCTACCGGGGTGACTTCGTCTTTGGGAGGACTTCA 720
Qy 1123 GGCCTGGCTGATCGAGATCAATTCACGCCCCCAGCCATGACCCCTCCACG 1172
Db |||||
721 NGCCCTGGCTGATCGAGATCAATTCACGCCCCCAGCCATGACCCCTCCACG 770

RESULT 6
AY415400 1265 bp DNA linear GSS 12-DEC-2003
LOCUS Mus musculus HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY415400
VERSION AY415400.1 GI:39771359
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1265)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
JOURNAL 14671302
PUBLISHED 2 (bases 1 to 1265)
REFERENCE Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
AUTHORS Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1265
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1265
/locus_tag="HCM5544"

gene
ORIGIN
Query Match 41.7%; Score 767; DB 9; Length 1265;
Best Local Similarity 78.6%; Pred. No. 5.6e-158;
Matches 949; Conservative 0; Mismatches 240; Indels 18; Gaps 2;

Qy 136 AAGACTTCCGGCGCACCATGCGATCCAGCATCTCAAGTGGTGGTCAGCCACGAGCT 195
Db |||||
1 ATGACTTCCGGCGCACCGTGGCTGCCAGCATCTCAAGTGGTGGTCTTCAACCAGAATT 60
Qy 196 GCAGCAGGACGACAGAGAACCCAGGACCCAGGAGGAGGAGGCGCGGAGCAGCGACC 255
Db |||||
61 ACT-----GCAGCAAAAGTCAAGGGCAAGAGTAGAAGAGGAGGAGGCGCAAGACAGTGACC 114
Qy 256 TGAGCAGCAGCAGAGATGCTGAAATGCTGAGGCAAAAGCTCAGGGCTCCCGGGCGCAGC 315
Db |||||
115 CCAGCCCCAAGAAAGATCTCGAATATCTGACCTCAAGCTACCAAGCCTCTCGGACAGG 174

```

```

Qy 316 TTGTGGATCATCGCTGCAAGGTGTGCCAGGCTTACCTTGGGGCAGCTGGAGCATGAGGACA 375
Db |||||
175 TTGTAGACACAGAGCTTGC AAGGTGTGCCAGGCTTACCTTGGGGCAGCTAGAGCATGAAGACA 234
Qy 376 TCAGACAGCTCAGCAGATGCGTGTGAGGACCTCACTGAGGCGGAGTGGAGGACCTGACCC 435
Db |||||
235 TAGACGTATCGGAGGCCAGCACCGAGGCTTGTCTGAGGAGGAATGAATGACCTTGACAC 294
Qy 436 AGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTC AAGAAATTACTTTT 495
Db |||||
295 AGCAGTACTACTTACTTGTTCATGGCAATGCTTCCCATCACTGATTCG A AAGTTACTTTG 354
Qy 496 CGCAGTCCAGGCTCTGCTGAATAGAAATACGCTGTGAAACCTTCAGACG GACATTTGACG 555
Db |||||
355 CGCAATGCCAAGCTCTGCTGAGTAAGATCAGTTCTGTGAAACCTCTCA GACAGAGATCGATG 414
Qy 556 GGCCTCGGGAACATCTGGATTATAAGCCCGGCGCAAGTCCCGGGCGG GAGACATAGTGT 615
Db |||||
415 GGATCCGGAACATCTGGATCATTAAGCTCGAGCAAGTCCCGGGGTG CAGATATTGTGT 474
Qy 616 GCATGGACCGTGTGGAGGAGATCCTTGAGCTGGCAGCTGCAGACCA C C C T C T T T C C A G G G 675
Db |||||
475 GTATGGACCGTGTGGAGAACATCCTGAGTCTGGTGGCCGCGAGAC G C C A G A A G G 534
Qy 676 ACAACAGTGGTGGTCCAGAAGTACATCAGAGCGCGCTGCTCATCTG TGAACACCAAGT 735
Db |||||
535 ACAACAAATGGTGGTCCAGAAGTACATCAGAGCCCCCAATGCTCAT C T A C G A C A C C A A G T 594
Qy 736 TCACACATCAGACAGTGGTTCCTCGTCAAGCTGGAGACCCCTTACC C A T C T G T T T C T A C A 795
Db |||||
595 TTGACATCAGACAGTGGTTCCTCGTCAAGCTGGAAATCCCTTAAC C A T C T G T T T C T A C A 654
Qy 796 AGGAGAGTTACTTTCGCGTCTCAACTCAGCGCTTCTCCCTGGACA G C T T G G A C A G C G C C A 855
Db |||||
655 AAGAAAGCTACCTCCGCTTCTCCACAGCGCTTCTCCCTGGACA G C T T G G A C A G C G C C A 714
Qy 856 TCACCTGTGCAACAACCGCGTCCAGAGTACCTGAAGAAATGATGTGG GCGCGCAGACCCCT 915
Db |||||
715 TCACCTGTGTAAACAATCCATCCAGAGCGCGCTCAAAAATGATAA A G A G C G C A G T C C G C 774
Qy 916 TGTGTCGCGCACACATGTGACAGCACCAAGTTCCAGAGGAGTACTC C A G C G C C A G G 975
Db |||||
775 TGTACTCTTGTCAATAACATGTGACCAAGCACTCGTTTCCAGGAG T A C C T T C G A A G A G G G 834
Qy 976 GCCTGTGCGCGTGTGGGGCAGCGTCACTTACCCGCTCCATGAAGA A G G C C A T C G C C C A C G 1035
Db |||||
835 GCCGAGGAGAAAGTGGGGTAGCATCATCTACCCGCTATGAA A A G A G C T G T C A C C A A T G 894
Qy 1036 CCATGAAGTGGCCCAAGGACCACTGAGAGCTTCGCAAGAACAGC T T T T G A G C T C T A C G G G G 1095
Db |||||
895 CCATGAGGGTAGCCAGGACCAAGTGAAGCCCGTAAGAACAGCTT C G A G C T C T A C G G A G 954
Qy 1096 CTGACTTCGCTCTGGAGGGAGCTTCAGGCGCTGGCTGATCGAGATC C A A T T C C A G C C C A 1155
Db |||||
955 CCGACTTCATCTCTGGGGCGAGACTTCAAGCCATGGCTTATCGAG A T C A C T C C A G C C C C A 1014
Qy 1156 CCATGCAACCCGTCACGCGCGGTCAAGCCCGCAGCTGTGTGCA C A G A G T G C A G A G G A C C A C C A 1215
Db |||||
1015 CCATGCAACCCCTCCACTCCCGTCAAGCTCAGCTCTGTG C C C A G A G T G C A G A G G A C C A C C A 1074
Qy 1216 TCAAGTGGCGG-----TGGACCGCAGCTGTGACATCGGCA A C T T C G G C A A C T T C G A G T C C 1263
Db |||||
1075 TCAAGGTGGTGGTGGATCGCAAACTGGACGAAACTGTGACATC G G C A A C T T T T G A G C T T C 1134
Qy 1264 TGTGGAGGCGCGGTGTTGAGCCGCCCTTACGCGGGTCCGAC C T C T G C G T G C G G G 1323
Db |||||
1135 TGTGGCGGCGAGCCTCGCGTGGAGCTGCCACCGTTCA A C G G G T C T G A C C C T C T G T G T G A A G 1194
Qy 1324 GCGTCAG 1330
Db |||||
1195 GTATCAG 1201

```

RESULT 7	775 bp	mRNA	linear	EST 18-OCT-2002
BU937372	775 bp	mRNA	linear	EST 18-OCT-2002
LOCUS	775 bp	mRNA	linear	EST 18-OCT-2002
DEFINITION	AGENCY: 10519180 NIH MGC 169 Mus musculus cDNA clone			
ACCESSION	IMAGE:6705256 5', mRNA sequence.			
VERSION	BU937372			
KEYWORDS	BU937372.1 GI:24126191			
SOURCE	EST.			
ORGANISM	Mus musculus (house mouse)			
REFERENCE	Mus musculus			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
JOURNAL	1 (bases 1 to 775)			
COMMENT	NIH-MGC http://mgs.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbe-remail.nih.gov Tissue Procurement: Dr. Jonathan Kuo, NIMH cDNA Library Preparation: Michael Brownstein Laboratory DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LUCM3002 row: 9 column: 16 High quality sequence stop: 569. Location/Qualifiers 1. 775 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clones="IMAGE:6705256" /lab_host="DH10B (TI-phage-resistant)" /clone_lib="NIH MGC 169" /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccctctggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCATGAGTATCAACGATGCGGATCGGCATCGCGGG-3' and 5'-ATTCTAGAGCGGCGGCGGATCGGATCGGATCGG-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."			
FEATURES				
source				
ORIGIN				
Query Match	24.1%	Score 443.6;	DB 5;	Length 775;
Best Local Similarity	74.5%	Pred. No. 7.7e-87;		
Matches	576;	Conservative 0;	Mismatches 185;	Indels 12; Gaps 1;
Qy	738	GACATCAGACAGTGGTTCTCGTCACGGATCGAACCCTCGACCTCTGTTCTTCTACAAG	797	
Db	1	GGCATCAGACAGTGGTTCTCGTCACAGACTGGAATCCCTTAACCATCTGTTCTACAAA	60	
Qy	798	GAGAGTTTACTTTCGGTCTTCAACTCAGCGCTTCTCCTCGACACAGCTGGACAGCGCCATC	857	
Db	61	GAAGCTACTCTCGCTTCTCCACACAGCGGTTCTCCTCGACACAGCTGGACAGCGCCATC	120	
Qy	858	CACCTGFGCAACAAACGCGCTGCGAGAGTACTGAAGAAATGATGTGGGCGCGACGCCATC	917	
Db	121	CACCTGTTGTAACAACTCCATCCAGAGCGGCTCAAAAATGATAAAGACGCGAGTCCGCTG	180	
Qy	918	CTGCGCGCACACACATGTGGACGACACAGGTTCCAGAGTACCTGACGCGCCAGGGC	977	
Db	181	CTACCTTGTTCATACATGTGGACGACGACCTCTGTTTCCAGAGTACCTGCGAGAGAGGGGC	240	
Qy	978	CGTGGCGCCGTGTGGGCGAGCGTCACTACCCCGTCCATGAAGAGGCGCATCGCCACGCC	1037	
Db	241	CGAGGAGGACGTGGGGTAGCATCATCTACCCGTCTATGAAGAGAGCTGTCCACCAATGCC	300	
Qy	1038	ATGAGTGTGCCAGGACACAGGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCT	1097	

```

/notes="hypothetical protein"
1..4184
/gene="DKFZp686D076"
2529..3587
/gene="DKFZp686D076"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH10554.1"
/db_xref="GI:50949619"
/translation="MDIDKLEAPLYLTPGWSLFLQRYQVYVHGEALRLHLDLTVQVR
CEDILQQLAVPQIDMEGRDNIWVKPGAKSRGIMCMHLEMLKLVNGPNVYMK
DKWVQVQTERPLIFGTGFDLQWPLVDWNPDLTWFYRDSYIRFSTQFSLKML
NSVHLNNSLOKLENSCHRPILLPPDNMSSORFOAHLQEMGAPNAWSTIIYVPGMKD
AVIHALTSQTVQCRKASPELYCADVFEGEDFQPLIEINASPTWAPSTAVTARLCA
GVQADTLRVVIDRMLDRNCTGAFELLYKQPVITSPASTPGPSCLLPMYSDTRRSSD
DSTASWWALRPCRQARP"

Query Match      22.9%; Score 421.4; DB 3; Length 4184;
Best Local Similarity 63.8%; Pred No. 7.5e-82;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;

Qy 232 GGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGCAGAGTGTGAAAATGCTGAGGCAA 291
Db 2391 GGGATCGTGACAGGCCCTCCCTATGTGACGAGACAAGCAGCCCAAGAACAGGAGAA 2450
Qy 292 AGCTCAG-GGGGCTCCGGGCGAGCTGTGTGACATCGCGGTGCAAGGTGTGCGAGGCCCTAC 350
Db 2451 AACCCAGTGTGTGTGTCGCCAGAGTTTGTGTGATGAAGCTCTGTGTGCGTGCAGGAGTAC 2510
Qy 351 CTGGGGCAGCTGGAGCATGAGGACATGCACACGTCAGCAGATGCGGTGGAGGACCTCACT 410
Db 2511 CTTAGCAACTTGTGGCCCATGATGACAGAGGACCTGGAGGCCCGCTGTACTCACC 2570
Qy 411 GAGGCCAGTGGGAGGACCTGACCCAGCAGTACTCTCCCTCGTTTCATGCGCATGCTTTC 470
Db 2571 CCCGAGGCTGTGCTTCTTCTCCAGGCTACTACCAAGTGTCCAGAGGGGAGAA 2630
Qy 471 ATCTCCAAATCAAGAAATTAATTTTTCGAGTGCCAGGCTCTGTGTAATAGAAATCACTCT 530
Db 2631 CTCAGGCACCTCGACACTCAGGTCAGGCTCGAGTGTGTGAGGACATCTCTGACAGCAGTGCAGGCC 2690
Qy 531 GTGAACCTCAGACGACATGAGGCTCGGAACATCTGGATTATTAAGACCGCGGCC 590
Db 2691 GTGTATCCCAAGATAGACATGGAAGGGGATCGCAACATCTGGATCGTGAAGCAGGAGGCC 2750
Qy 591 AAGTCCCGGGCCGAGACATAGTGTGATGACCGTGTGGAGGAGATCTCTGGAGCTGGCA 650
Db 2751 AAGTCCCGGGCAGGACATCATGTGATGGACCACTCTGGAGGAGATGCTGAAGCTGGT 2810
Qy 651 GCTGCAGACCACTCTTTTCCAGGGAACAAGTGGGTGTGTCAGAAAGTACATCGAGACG 710
Db 2811 AACGGCAACCCGTGTGTGATGAAGGACGGCAAGTGGGTGTGTCAGAAAGTATATTGACGG 2870
Qy 711 CCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGTTCCTCGTCAAGGACTGG 770
Db 2871 CCCCCTCTCATCTTTTGGCAACAAGTTTGAACCTCAGACAGTGTTCCTGGTAACTGACTGG 2930
Qy 771 AACCCCTGACCATCTGCTTCTCAAGGAGAGTTACTTGGGTTTCTCAACTCAGCGGCTTC 830
Db 2931 AACCCACTTACCTGTGTGTCTACCGCGACAGCTATATCGCTTTTCCAGCAGCCCTTC 2990
Qy 831 TCCCTGACCAAGCTGGACAGCGCCATCCACTGTGCAACAACCGCGTCCAGAAAGTACTCG 890
Db 2991 TCCCTGAAGAACCTGGACAATCAGTGCACCTGTGCAACAACCTCATCCAGAGACACCTG 3050
Qy 891 AAGNATGATGTGGCGCGAGCCCTGCTGTCGCGCACACACATGTTGGACGACGACACAGG 950
Db 3051 GAGAACTCATGTGCCATCGGCATCCACTGCTTCCGCGCAGACAACATGTGTGTGTAGCCAGAG 3110
Qy 951 TTCCAGGAGTACCTGACGAGCGCCAGGCGGTGGCGCGTGTGGGCGCAGCGTCACTACCCG 1010
Db 3111 TTCAGGCCCCACCTGACGAGGAGATGGGTGGCCCAATGCTTGTGCCACCATCATCGTGCCT 3170
```

ORIGIN

```

Qy 1011 TCCATGAGAGGCCCATCGCCACGACCATGAAGTGGCCAGCAGCAGCTGGAGCCTCGC 1070
Db 3171 GGCATGAAGGATGCTGTGATCCACGCACATTCAGACCTCCAGGACACACCGTGCAGTGTGCG 3230
Qy 1071 AAGAAACAGCTTTTCTGAGCTTACGGGGGTGACTTCGTCTCTTGGGAGGAGACTTCCAGGCCCTGG 1130
Db 3231 AAGCCAGCTTTGAGCTCTATGGCGTGTGACTTCGTGTTCGGGGAGGACTTCCAGGCCCTGG 3290
Qy 1131 CTGATCGAGATCAATTCACGCCCCACCATGCACCCGTCACCGCGGTTCACGCGCCAGCTG 1190
Db 3291 CTGATTCAGATCAACGCGCAGCCCAAGATGGCACCTCCACAGCAGTCACTTCCGCGGCTC 3350
Qy 1191 TGTGCACAGTGTGAGGAGGACACATCAAGGTGGCC-----GTGGACCGCAGC 1238
Db 3351 TGTGCTGGCGTGAAGCTGACACCTCGCGGTGTTCATTGACCGGATGCTGGACCGCAAC 3410
Qy 1239 TGTGACATCGCAACTTCGAGCTCTCTGTGGAGGAGCGCGGT 1279
Db 3411 TGTGACACAGGAGCCTTTGAGCTCATCTATAAGCAGCCCGT 3451
```

```

RESULT 9
CK597795 777 bp mRNA linear EST 22-JAN-2004
AGENCOURT 17837841 NIH_MGC_238 Rattus norvegicus cDNA clone
IMAGE:7135519 5', mRNA sequence.
CK597795
CK597795.1 GI:41110910
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
```

```

REFERENCE
1 (bases 1 to 777)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM15042 row: 0 column: 05
High quality sequence stop: 693.
```

```

FEATURES
Location/Qualifiers
1..777
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7135519"
/tissue_type="testis, pooled"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_238"
/notes="Organ: testis; Vector: pExpress-1; Site: 1: EcorV;
Site 2: NotI; RNA obtained from testis tissue of 8 wk old
animal. Tissues were snap-frozen and kept at -80C before
RNA extraction and purification (Tri-reagent method). CDNA
was primed using oligo-dT primer:
5'-pGACTAGTCTAGATCGAGCGCCGCCCTTC-3' and cloned into
the EcorV/NotI sites of pExpress-1. Size-selection >1.4kb
resulted in an average insert size of 1.9 kb. This primary
library is normalized (non-normalized primary library is
NIH_MGC_237) and was constructed by Express Genomics
(Frederick, MD)"
```



```
/clone="A630053H17"
/tissue type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
84..2198
CDS
/notes="unnamed protein product; HOTT1 PROTEIN homolog
[ Homo sapiens ] (spt8) [ Q9UI99, evidence: FASTV, 85.8%ID,
93.4%length, match=821 ]
putative"
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"
/translations="MDTSSPFAAGQGEASPPSLVSISSPQLSRMWRNETPVLWT
TREVDLDCRSLSDQMLNHLVARAGSTTKVGLCLNLENLWPFDEADADSPFPCYRLG
AEDDKAFIEDFWLTAARNVLKLVKLEKSQISIQAREEAPEDTPKPKKLVLT
VSSDFVDEALSCOEHLSSIAHKDIKDQNSPLYLSPDDWSQFLQRYQIVHGEABL
YLSDFQVQEDIIQQLNVVLPQDMDGRDNIWIKPKAKSRGRGIMCMRLDMLKL
DCNPMKMDGKWTVKYIERPLLI FGTKFLRQWLFTDWNPLTVWPFYRDSYIRFSTQ
PFSLKNDLVHLCNNSIORHLEASCHRRPMLPDNNWSSORFOALHVEDAPKANS
VIVPGMKAAVIHALQTSODNVOCRKASFEYLGFADVFGEDEFQWLEINASPTAPST
AVTARLCAGVQADTLRVIRDLRDCDTSQAFELIYKQPAVEVPQVIRLLVGGSTI
KRPVVGHRHTVRSLSLPHLLTQGGSGKDSQSPHRSASRNKRAESLEHTEKEPEP
AAVASGSGKGAAPFPFSLHSAWLPSPRVHRPQGRVLRLOHDLQVSGKALSTTKGA
LMTLPKLVMSPPPHDLKAPSLMKPKGVGFELCCTTWRVVLSSGIGBEGRQRAA
PRPSAPGKGLSSTPCSTET"
```

ORIGIN

Query Match 21.7%; Score 399; DB 3; Length 2520;
Best Local Similarity 61.8%; Pred. No. 5.9e-77;
Matches 658; Conservative 0; Mismatches 395; Indels 12; Gaps 1;

```
QY 291 AAGCTAGGGGGCTCCGGGGCAGCTTGTGCAATCGCGTCAAGGTGTCAGGCGCTAC 350
DB 552 AAGCTAGTACCGGTGCTCCTCAGACTTTGTGCGATGAGGCTCTGATGCGATCCAGGAGCAC 611
QY 351 CTGGGGCAGCTGAGCATGAGGACATCGACACGTCAGCAGATCGCGTGGAGGACCTCACT 410
DB 612 CTTAGCAGCATAGCCCAAGAAGCATTGACAGAGACCCCAACTCCCGCTCTACTCTCAGC 671
QY 411 GAGCGCAGTGGAGGACCTGACCCAGCAGTACTTCTCCCTCGTTATGCGGAGTCTTTC 470
DB 672 CCGGATGATTGGTCCAGTTCCTTTCAGCGCTACTTACCAGATAGTTTATGAAAGGGCAGAG 731
QY 471 ATCTCCAAATCAAGAAATTAATTTTTCAGTGCAGGCTCTGCTGATAGATACAGTCT 530
DB 732 CTGAGTACTGAGGTGCAAGTCCAGCGCTGTGAAGACATCTCTACAGCAGTCTTCAAGAC 791
QY 531 GTCAACCTTCAGACGGACATTTAGCGGCTCCGGAACATCTGGATTATAAGCCCGCGGCC 590
DB 792 GTGGTACCCAGTTAGACATGGAGGGGATCGGAACATCTGGATCGTGAAGCCCGGAGCC 851
QY 591 AAGTCCCGGGCCGAGACATAGTGTGATGAGACCGTGTGAGGAGATCTCTGGAGCTGGCA 650
DB 852 AAGTCCCGGCGCGAGGATTATGTGATCAACCGCTGATGAGATGCTGTAAGCTGGTG 911
QY 651 GCTGACAGACCACTCTTTCAGGAGCAACAAGTGGTGTGTCAGAGTACATCGAGAG 710
DB 912 GACTGCAACCCCATGCTCATGAAGGATGGCAAGTGGATCGTGCAGAGTACATTTGAGCGG 971
QY 711 CCGCTGCTCATCTGTGACACAAAGTTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGG 770
DB 972 CCGCTGCTCATCTGTGACACAAAGTTTCGACATCAGACAGTGGTTCCTCGTCGAGTACTGG 1031
QY 771 AACCCCTGACCAATCTGGTTTCTAAGGAGAGTACTTGGCGTTTCTCACTCAGCGCTTC 830
DB 1032 AACCCCTGACCGTGTGGTTTCTACCGGAGACAGTACTATTCGCTTCTCCACACAGCCTTC 1091
QY 831 TCCTTGACAAAGCTGACAGCGGCATCCACTGTGCAACAAACCGCGTCTCGAAGTACTCTG 890
DB 1092 TCCTTGAAAGAACCTGGCAACTCTGTGCAACTCTGTGAACAACTCCATCCACAGACACTTG 1151
QY 891 AAGAAATGATGTGGCGCGACCCCTGCTGCGGACACACAAATGTGGACCAACACAGG 950
```

```
DB 1152 GAGGCTTCTGTCAACCGGACCCGATGCTGCCCCCAGATAAATGTGGTCCAGCAGAGG 1211
QY 951 TTTCCAGAGTACTCTGAGAGCCCGAGGGCCGTGGCGCCTGTGGGGCAGCGTCACTATCCCG 1010
DB 1212 TTTCCAGGCCCACTTGCAGGAGGTAGATGCCCCCAAAGCCTGTGTCACGCGTCACTGTCGA 1271
QY 1011 TCCATGAAGAAGCCATCGCCACGCGCATGAAGTGTGCCAGGACACACGCTGGAGCCTCGC 1070
DB 1272 GGCATGAAGGCTGCTGTGATCCATGCGCTCGACACCTCCCAAGATAACGTGCGAGTCCGG 1331
QY 1071 AAGAACAGCTTTTGTGCTTACGCGGCTGACTTCTGCTCTTGGGAGGAGTTCAGGCCCCCTGG 1130
DB 1332 AAGGCCAGCTTTTGTGCTTATGCGGCGAGACTTTGTGTTTGGGGAAGACTTTCAGGCCCTGG 1391
QY 1131 CTGATCAGATCAATTCAGCCCCCAGCATGCAACCGCTGCAAGCGGTGACGCGCCAGCTG 1190
DB 1392 TTTGATGAATCAATGCCAGCCCCCAGCATGCGACCTTTCACGCGTGTGCTGCGCGCTC 1451
QY 1191 TGTGACAGGTGACAGAGACAC-----CATCAAGTGTGGCGTGGACCGCAGC 1238
DB 1452 TGTGCGGTGTGCGAGCAGATACCTTGTGCTGATGCGCGCTTGGACCGCAGC 1511
QY 1239 TGTGACATCGGCAATTCAGGCTCTTGTGAGGACGCGGTGTTGAGCGCCCCCAATTC 1298
DB 1512 TGTGATACGGAGCCTTTTGAGCTCATCTATAAGCAGCCTGCTGTGAGGTGCTCAGTAC 1571
QY 1299 AGCGGTGCGACCTTGTGCGGCGGTGCTGAGGAGAGCC 1343
DB 1572 GTGGGTATCGGCTCTAGTGGAGGCTCTTACCATCAAGAAGCCC 1616
```

RESULT 11

BM808637

LOCUS

DEFINITION

5', mRNA

1023 bp

mRNA

linear

EST 05-MAR-2002

AGENCOURT 6617786 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734232

5', mRNA sequence.

BM808637

BM808637.1

GI:19125460

EST.

Homo sapiens

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1023)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Invitrogen

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM12739

row: d

column: 09

High quality sequence start: 6

High quality sequence stop: 664.

Location/Qualifiers

1..1023

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5734232"

/tissue_type="hippocampus"

/lab_host="DH10B"

/clone_lib="NIH_MGC_124"

/note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: EcorV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is

FEATURES

source

normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

ORIGIN		Query Match		20.6%; Score 378.4; DB 4; Length 1023;		Best Local Similarity		69.5%; Pred. No. 1.8e-72;		Matches 514; Conservative 0; Mismatches 226; Indels 0; Gaps 0;	
QY	497	GCAGTCCAGGCTCTGCTGAATAGATCAGCTGTGTAACCCCTCAGACGGACATTGACGG	556			Db	49	GGCTGTGAGGACATCTCGACAGCTGCGAGCGGTGGTACCCAGATAGACATGGAAGG	108		
QY	557	GCTCCGGAACATCTGATTATAAAGCCGCGCAAGTCCCGGGCCGAGACATAGTGTG	616			Db	109	GGATGCAACATCTGATCGTGAAGCCAGGAGCAAGTCCCGGAGGAGGCATCATGTG	168		
QY	617	CATGGACCGTGTGGAGGAGATCTCGAGCTGGCAGCTGCAGACCACTCTTTTCCAGGA	676			Db	169	CATGGACCACTGGAGGAGATGCTGAAGCTGTGTAACGCAACCCCGTGTGATGAAGGA	228		
QY	677	CAACAGTGGTGGTCCAGAGTACATCGAGCGCGCTGCTCATCTGTGACACCAAGTT	736			Db	229	CGGCAAGTGGTGGTGCAGAGTATATTGAGCGGCCCTCTCATCTTTTGGCACCAGTT	288		
QY	737	CGACATCAGACAGTGGTCTCTGCTCAGGACTGGAAACCCCTGACCATCTGGTTCTACAA	796			Db	289	TGACCTCAGACAGTGGTCTCTGCTAATGACTGACTGGAACCACTTACCGTGTGTTACCG	348		
QY	797	GGAGTTACTTGGGTTCTCAACTCAGCGTCTTCCCTGGACAGCTGGACAGCGCCAT	856			Db	349	CGACAGTATATCCGCTTTTTCACGCGAGCCCTTCTCCCTGAAAGAACCTGGACAACTCA	408		
QY	857	CCACCTGTGCAACACGCGCTCCAGAGTACCTGAAGATGATGTGGCGCGCAGCCCCCT	916			Db	409	GCACCTGTGCAACACTCCATCAGAGCACCTTGGAGAACTCATGCGATCGGCATCACT	468		
QY	917	GCTGCCCGCACCAACATGTGGACAGCACCAGGTTTCCAGGAGTACTCTCAGCGCCAGG	976			Db	469	GCTTCCGCGCAGACAACTATGTGTTAGCCAGAGTTTCCAGGCCCACTTCGAGGATGG	528		
QY	977	CGTGCGCGCGTGTGGGCGAGGTATCTACCGCTGCATGAAGAGGCCATCGCCACGC	1036			Db	529	TGCCCCAAATGCTTGGTCCACCATCATCTGCTGCTGCAATGAAGGATGCTGTATCCAC	588		
QY	1037	CATGAAGTGGCCAGACACAGTGGAGCTCGCAAGACAGCTTTGAGCTCTAGGGGC	1096			Db	589	ACTTCAAGCTCCAGGACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTATGGCGC	648		
QY	1097	TGACTTCTGCTTGGAGGGACTTCAGGCCCTCGCTGTGATCGAGATCAATTCAGGCCCA	1156			Db	649	TGACTTCTGCTTGGAGGGAGCTTCAGGCCCTCGCTGTGATGAGATCAACGCCAGCCCC	708		
QY	1157	CATGACCCGTCACCGCGTCAAGCCAGCTGTGTGACAGGTGTCAGGAGGACACCAT	1216			Db	709	GATGGCACCTCCACAGCACTCACTCCCGGCTCTGTGCTGGCTGCAAGCTGCACACCT	768		
QY	1217	CAAGTGGCGGAGCCGCA	1236			Db	769	GGCGTGGTCAATTGACCGGA	788		
RESULT 12											
BM808516											
LOCUS											
DEFINITION											
AGENCOURT_6617783 NIH_MGC_124 Homo sapiens cdna clone IMAGE:5734160											
5', mRNA sequence.											
ACCESSION											
BM808516											
VERSION											
BM808516.1 GI:19125339											
KEYWORDS											
EST.											
SOURCE											
Homo sapiens (human)											
ORGANISM											
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;											

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1022)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cdna Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12739 row: a column: 09
High quality sequence start: 78
High quality sequence stop: 658.

Location/Qualifiers
1..1022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734160"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

FEATURES
source
1..1022
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734160"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

ORIGIN
Query Match 19.6%; Score 360; DB 4; Length 1022;
Best Local Similarity 68.8%; Pred. No. 2e-68;
Matches 495; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 497 GCAGTCCAGGCTCTGCTGAATAGATCAGCTGTGTAACCCCTCAGACGGACATTGACGG 556
Db 40 GCGCTGTGAGGACATCTCTGTCAGCAGCTGCAGGCCGTGGTATCCAGATAGACATGGAAG 99
QY 557 GCTCCGGAACATCTGATTATAAAGCCGCGCAAGTCCCGGGCCGAGACATAGTGTG 616
Db 100 GGATGCAACATCTGATCGTGAAGCCAGGAGCAGTCCCGGAGGAGGCATCATGTG 159
QY 617 CATGACCCGTCACCGCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTTCCAGGA 676
Db 160 CATGACCACTTGGAGGAGATGCTGAAGCTGGTGAACCGCAACCCCGTGTGATGAAGGA 219
QY 677 CAAAGTGGTGGTCCAGAGTACATCGAGAGCGCGTGTCTCATCTGTGACACCAAGTT 736
Db 220 CGGCAAGTGGTGGTGCAGAGTATATTGAGCGGCCCTCTCATCTTTTGGCACCAGTT 279
QY 737 CGACATCAGACAGTGGTCTCTGCTCAGCGACTGGAAACCCCTGACCATCTGGTTCTACAA 796
Db 280 TGACCTCTTACAGTGGTTCCTGTTAACTGACTGGAACCACTTACCGTGGTGTCTACCG 339
QY 797 GGAGAGTTACTTGGCGGTTCTCAACTCAGCGTTCCTCGTGGACAAAGCTTCCCTGGACAAAGTGGAGGCCCAT 856
Db 340 CGACAGCTATATCCGCTTTTTCACGCGAGCCCTTCTCCCTGAAAGAACCTGGACAACTCACT 399
QY 857 CCACCTGTGCAACACGCGTCCAGAGTACCTGAAGATGATGTGGCGCGCAGCCCCCT 916
Db 400 GCACCTGTGCAACACTCCATCAGAGCACCTTGGAGAACTCATGCGATCGGCATCACT 459
QY 917 GCTGCCCGCACCAACATGTGGACAGCACCAGGTTTCCAGGAGTACCTTCAGCGCCAGG 976
Db 460 GTTTCGCGCAGACACATGTGTCTTAGCCAGAGGTTCCAGGCCACCTCGCAGGAGATGG 519
QY 977 CCGTGGCGCGGTGTGGGCGAGCGTCTATCTTACCGTTCATGAAGAGGCCATCGCCACGC 1036

Db 520 TGCCCCAAATGCTTGGTCCACCATCATGTCCTGGCATGAGATGCTGTATCCAGC 579
 Qy 1037 CATGAAGTGGCCAGGACCACTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGC 1096
 Db 580 ACTTCAGACCTCCAGGACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTATGGCGC 639
 Qy 1097 TGACTTCGCTCTGGGAGGACTTCAGGCCCTGGCTGATCGAGTCAATTCACGCCCCAC 1156
 Db 640 TGACTTCGCTCTGGGAGGACTTCAGGCCCTGGCTGATCGAGTCAATTCACGCCCCAC 699
 Qy 1157 CATGACCCCGTCCACGCCGGTTCAGGCCCTGGCTGATCGAGTCAATTCACGCCCCAC 1216
 Db 700 GATGGCACCCCTACAGCAGTCACTGGCGGCTCTGTGCTGGCTGCAAGTCAACCCCT 759

RESULT 13
 BM717061
 LOCUS
 DEFINITION UI-E-EJ0-ahk-e-05-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-ahk-e-05-0-UI 5', mRNA sequence.
 BM717061
 ACCESSION BM717061.1 GI:19030319
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 693)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Rresearch
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES
 source
 1..693
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-ahk-e-05-0-UI"
 /tissue_types="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies)" (T1 phage resistant)"
 /clone_lib="UI-E-EJ0"
 /notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT7T3-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes,

AGATCAAGA; lens, CGATTAGCGA; eye anterior segment,
 AATGCCCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina
 Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
 library was created for the program, Gene discovery in the
 Visual System, supported by National Eye Institute (NEI)."

ORIGIN
 Query Match 17.5%; Score 321; DB 4; Length 693;
 Best Local Similarity 69.1%; Pred. No. 7.2e-60;
 Matches 473; Conservative 0; Mismatches 200; Indels 12; Gaps 2;
 Qy 603 CGAGACATAGTGTGATGACCGGTGGAGAGATCTCGAGCTGGCAGCTGCAGACAC 662
 Db 1 CGAGGCATCATGTGATGACACCTGGAGAGATGCTGAAGCTGGT-GAAGCGACCCC 59
 Qy 663 CCTCTTTCCAGGGACAAACAAGTGGTGGTCCAGAAGTACATCGAGACCGCTCTCATC 722
 Db 60 GTGTGATGAAGACGCGCAAGTGGTGGTGCAGAAGTATATTGAGCGCCCCCTCTCATC 119
 Qy 723 TGTGACACCAAGTTGACATCAGACAGTGTTCCTGTCAGGACTGGAAACCCCTGACC 782
 Db 120 TTTGGCACCAGTTTGACCTCAGACAGTGTTCCTGGTAACTGACTGGAAACCCATTACC 179
 Qy 783 ATCTGTTTCTACAAGAGAGTACTTTCGCGTTCCTCAACTCAGGCTTCTCCCTGCAAG 842
 Db 180 GTGTGTTTACCGGACACCTATATCCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAAC 239
 Qy 843 CTGGACAGCGCCATCCACCTGTGCAACAACCGCTGCCAGAAGTACCTTGAAGAATGATGTG 902
 Db 240 CTGACAACTCAGTCACCTGTGCAACAACCTCCATCCAGAAGCACCTGGAGAATCATGC 299
 Qy 903 GGGCGAGCCCCCTGTGCGCCGACACAACATGTGGACACGACCAAGTTCAGAGGTAC 962
 Db 300 CATCGGCATCCACTGTCTCCGCCAGACAACATGTGGTCTAGCCAGAGAGTTCCAGGCCAC 359
 Qy 963 CTGACGCGCAGGCGCTGGCGCGTGTGGGAGAGTCTATCTACCGTCCATCAAGAAG 1022
 Db 360 CTCGAGAGATGGTGGTGGCCCAATGCTTGTTCACCATCATCTGCTGGCATGAAGAT 419
 Qy 1023 GCATCGCCACCGCCATGAAGTGGCGCCAGGACCACTGGAGGCTCCCAAGAACAGCTTT 1082
 Db 420 GCTGTGATCCAGCACTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTT 479
 Qy 1083 GAGCTCTAGGGGCTGACTTGTGCTTGGAGGAGACTTCAGGCCCTGGCTGATCGAGATC 1142
 Db 480 GAGCTCTATGGCGCTGACTTGTGTTGGGGAGGAGCTTCCAGGCCCTGGCTGATGAGATC 539
 Qy 1143 AATTCCAGCCCCACCATGACCCGCTCCAGCGGTCCAGCCCGCAGCTGTGTGCACAGGTG 1202
 Db 540 AAGCCAGCCCCACGATGGCACCTCCACAGCAGTCACTGCCCCGGCTCTGTGTGGCGTG 599
 Qy 1203 CAGGAGGACAC-----CATCAAGTGGCGCTGGAGCCGAGCTGTGACATCGGCA 1251
 Db 600 CAAGCTGAACCTCGCGGTGCTATTGACCGGATGTGGACCGCAGTGTGACACAGGAG 659
 Qy 1252 ACTTCGAGCTCTGTGGAGGACGCC 1276
 Db 660 CCTTTGAGCTCATCTATAAGCAGGC 684

RESULT 14
 CK653583
 LOCUS
 DEFINITION AGENCOURT 17673480 NIH_MGC_237 Rattus norvegicus cDNA clone
 IMAGE:7113394 5', mRNA sequence.
 CK653583
 ACCESSION CK653583.1 GI:41389106
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Rattus norvegicus (Norway rat)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 861)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Howard Jacobs
 cDNA Library Prepared by: Express Genomics
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM14985 row: e column: 08
 High quality sequence stop: 304.
 Location/Qualifiers
 1..861
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="IMAGE:7113394"
 /tissue_type="testis, pooled"
 /lab_host="DH10B TONa"
 /clone_lib="NIH MGC 237"
 /note="Organ: Testis; Vector: pExpress-1; Site: 1: EcoRV; Site 2: NotI; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer:
 5'-pGACTAGTCTAGATCGGAGCGGCCCT(7)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb resulted in an average insert size of 2.4 kb. This primary library is not normalized (normalized primary library is NIH_MGC_238) and was constructed by Express Genomics (Frederick, MD)"

ORIGIN
 Query Match 17.2%; Score 316; DB 7; Length 861;
 Best Local Similarity 74.2%; Pred. No. 9.4e-59;
 Matches 439; Conservative 0; Mismatches 150; Indels 3; Gaps 3;
 QY 595 CCGGGGCGGACATAGTGTGATGACCGTGTGGAGAGATCTCTGGAGCTGGCAGCTG 654
 DB 12 CCGGATTGGAGATATTGTGTGATGACCGTGTGGAGAACATCTCTGGATCTGGTGGCC- 70
 QY 655 CAGACACCTCTTTCCAGGGACAAAGTGGGTGTCAGAGTACATCGAGACGCCG 714
 DB 71 CAGACAGGCGAGACCACGAGGACAAATGGGTGTCATAGTACATCGAGACCCCGA 130
 QY 715 TCCTCATCTGTGACACCAAGTTCGATCAGACAGTGGTTCCTCGTCACGGACTGGAACC 774
 DB 131 TCCTCATCTGTGACACCAAGTTCGATCAGACAGTGGTTCCTGTGTCACAGACTGGAATC 190
 QY 775 CCCTGACCATCTGGTTCACAGGAGTACTTGGGGTTCCTCAACTCAGCGCTTCTCCC 834
 DB 191 CCTTAACCATCTGGTTCACAAAGAGAGTACTCTCGGTTCTCCAGCAGCGCTTCTCCC 250
 QY 835 TGGACAGCTGTGACAGCGCCATCCCTGTGCAACACACCGCTCCAGAGTACCTGGAAGA 894
 DB 251 TGGACAAATGTGACAGCGCTTCCACTGTGTAAACAACTCCGTCCAGAGCGGTGGCATAA 310
 QY 895 ATGATGTGGGCGCAGCCCTCTGTCGCCGACACAAATGTGGACACACAGCTTCC 954
 DB 311 ATGACAAATAAGAGTCCACTGTCTACCTTGCATTAACATGTGGACCATCACCGCTTCC 370
 QY 955 AGGAGTACCTGTGACGCCAGGCGCGTGGCCGCTGTGGGGCAGCGCTCATCTACCCGTCCA 1014
 DB 371 ATGAGTACCTGTGATATATATGGGCGGAGGATGCACATGGGGTAGTATCATCTACCCGTCTA 430
 QY 1015 TGAAGAGGCCATCGC-CCACGCCATGAAGGTGGGCCCGACGACCATCGTGGAGCCTCGCAAG 1073

Db 431 TGAAGAAGAGCTGCTCACTCAACGTCAATGATGGTGGCCAGATCTTGTGAAGCCCGTAAG 490
 QY 1074 AACAGCTTTGAGCTCTACGGGCTGACTTCGTCCTTGGAGGAGCTT-CAGGCCCTGGCT 1132
 Db 491 ATTATCTTCGATCTCTATGGAGCTGACTTTAATCTGGGCGGTGACTTATAGTCTCTGTGT 550
 QY 1133 GATCGAGATCAATTCAGCCCCCACCATGACCGCTCCAGCCGGTCACGCC 1184
 Db 551 CTTCAAGATCAACTTCTACTCTGCCTTGCAGCTCCTACTTTTGTAAAAACC 602

RESULT 15
 CD515431 896 bp mRNA linear EST 06-JUN-2003
 LOCUS AGENCOURT 14364765 NIH_MGC.181 Homo sapiens cDNA clone
 DEFINITION IMAGE:30407505 5', mRNA sequence.
 ACCESSION CD515431
 VERSION CD515431.1 GI:314447149
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 896)
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Daniela S. Gerhard, Ph.D.
 COMMENT Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-r@mail.nih.gov
 cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: NDAM488 row: p column: 10
 High quality sequence stop: 678.
 Location/Qualifiers
 1..896
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30407505"
 /tissue_type="White Matter"
 /dev_stage="Unknown"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH MGC 181"
 /note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.42 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 16.7%; Score 307.6; DB 6; Length 896;
 Best Local Similarity 70.0%; Pred. No. 6.6e-57;
 Matches 434; Conservative 0; Mismatches 174; Indels 12; Gaps 1;
 QY 672 AGGGACAAACAGTGGGTGGTCCAGAGTACATCGAGACCGCGTGTCTATCTGTGACACC 731
 Db 25 AAGGACGGCAAGTGGGTGGTGGAGATATTTGAGCGGCCCTCTCATCTTTGGCACC 84
 QY 732 AAGTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGGAACCCCTGACCATCTGGTTC 791
 Db 85 AAGTTTGACCTCAGACAGTGGTTCCTCGTGAACCTTACCTGGAACCCCACTTACCGTGTGTTTC 144
 QY 792 TACAAGGAGAGTTACTTGGGTTTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGC 851

```
Db 145 TACGGACAGCTATATCGCTTTCCACGAGCCCTTCTCCCTGAGAACCTGGACAAC 204
Qy 852 GCCATCCACCTGTGCAACAACGGCGTCCAGAATGACCTGAAGATGATGTGGGCGCGCAGC 911
Db 205 TCAGTGACCTGTGCAACAACCTCCATCCAGAAGCACCTGGAGAACTCATGCCATCGGCAT 264
Qy 912 CCCTGTGCTGCCCGCACACAACATGTGGACAGACACAGGTTCCAGGAGTACCTGCGACGCG 971
Db 265 CCAGTGTTCGGCCAGACACATGTGTCTAGCCAGAGGTTCCAGGCCACCTGCGAGGAG 324
Qy 972 CAGGGCGGTGGCGCGGTGTGGGCGAGCGTCACTTACCCGTCCATGAAGAAGGCCATCGCC 1031
Db 325 ATGGGTGCCCAAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGGATGCTGTGATC 384
Qy 1032 CAGCCCATGAAGTGGCCAGGACCAAGTGGAGCGCTCGCAAGAACAGCTTTGAGCTCTAC 1091
Db 385 CACGCACTTCAGACCTCCCAAGACACCGTGCAATGTGGAAGGCCAGCTTTGAGCTCTAT 444
Qy 1092 GGGGCTGACTTTCCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGC 1151
Db 445 GGGCTGACTTTCGTTCGGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAACGCCAGC 504
Qy 1152 CCCACCATGACCCGTCACGCGGTGACGCGCCAGCTGTGTGCACAGGTGCGAGGAGGAC 1211
Db 505 CCCACGATGGCACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGCAAGCTGAC 564
Qy 1212 ACCATCAAGTGGCC-----GTGGACCGCAGCTGTGACATCGGCAACTTTCGAG 1259
Db 565 ACCCTGGCGTGTGTCATTGACCGGAGGCTGGACCGCAACTGTGACACAGGAGCCTTTGAG 624
Qy 1260 CTCCTGTGGAGGCGAGCGGT 1279
Db 625 CTCATCTATAAGCAGCCCGT 644
```

Search completed: September 25, 2005, 01:55:59
Job time : 7070.48 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 01:51:04 ; Search time 1115.21 Seconds
(without alignments)
9756.464 Million cell updates/sec

Title: US-10-615-659-1

Perfect score: 1838

Sequence: 1 tggagcagccctgggcccc.....actctccaaggcggaattc 1838

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1838	100.0	1838	12	ADJ93357 Human BGS
2	1782.8	97.0	3554	12	ADJ93364 Human tub
3	1678.8	91.3	3465	12	ADJ93363 Human BGS
4	1677.2	91.3	1939	12	ADJ93361 Human BGS
5	1574.8	85.7	1859	12	ADJ93362 Human BGS
6	726	39.5	726	12	ADJ93359 Human BGS
7	501	27.3	101270	12	ADQ17814 Human sof
8	432.8	23.5	2380	6	AAS9894 Polynucle
9	431.6	23.5	2326	4	AAL16735 Human cDN
10	427	23.2	490	12	ACH91699 Human gen
11	421.4	22.9	2538	7	ADJ73168 Human kid
12	421.4	22.9	2553	13	ACN37881 Tumour-as
13	421.4	22.9	2881	3	AAC77214 Human ORF
14	421.4	22.9	3001	4	AAL58606 Human pol
15	421.4	22.9	3001	5	ADQ98824 DNA encod
16	421.4	22.9	3001	9	ADB48584 Novel hum
17	419.8	22.8	3294	8	ABX34502 Human mdd
18	418	22.7	2848	11	ADM03081 Human cDN
19	413	22.5	2611	3	AAF21812 Human bre
20	347.4	18.9	2412	12	ADH45468 Human mol

21	276	15.0	2469	12	ADH45478 Human mol
22	272.4	14.8	2241	12	ADJ93378 Human BGS
23	231.4	12.6	5282	12	ADQ64426 Novel hum
24	231	12.6	2044	4	AAI60392 Human pol
c	224.6	12.2	3828	13	ADR07582 Full leng
26	213	11.6	2979	4	ABL17657 Drosophil
27	206.8	11.3	2217	10	ACA92443 DNA encod
28	192.4	10.5	1567	8	ACC46592 Human dit
29	190.6	10.4	1958	4	AAL18689 Human cDN
c	190.6	10.4	10468	4	AAK77961 Human imm
31	174.2	9.5	2543	4	ABL17449 Drosophil
32	166	9.0	1728	5	AAS70158 DNA encod
33	164.4	8.9	2250	5	AAS69383 DNA encod
34	163.4	8.9	418	8	ABX52103 Bovine ES
35	147.8	8.0	1085	4	AAF63819 Human sec
36	144.8	7.9	4615	4	ABL17448 Drosophil
c	136.8	7.4	5728	4	ABL17656 Drosophil
38	131.8	7.2	492	9	ACH48041 Human lun
c	130	7.1	525	12	ACH77979 Human gen
40	92.2	5.0	207	6	ABS69387 Novel mur
c	80	4.4	80	12	ADJ93402 Human BGS
42	79.4	4.3	1571	4	ABL14515 Drosophil
43	72.8	4.0	996	10	ADC32299 Human nov
44	72.8	4.0	2259	12	ADQ64293 Novel hum
45	72.8	4.0	2417	10	ADC30440 Human nov

ALIGNMENTS

RESULT 1
ADJ93357
ID ADJ93357 standard; cDNA; 1838 BP.
XX

AC ADJ93357;

DT 06-MAY-2004 (first entry)

XX Human BGS-42 cDNA sequence SeqID1.

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nontropic; antiparkinsonian; antiarthritic; antiseborrheic;
anti-Hiv; antibacterial; immunosuppressive; antitubercular;
dermatological; tyrosine kinase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.

XX Homo sapiens.

Key Location/Qualifiers
CDS 153..1778
/*tag= a
/product= "Human BGS-42 protein"

XX WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu, S, Nelson TC;

XX WPI; 2004-099381/10.
DR P-PSDB; ADJ93358.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
PS Claim 1; SEQ ID NO 1; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antidiabetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.
XX
SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Query Match 100.0%; Score 1838; DB 12; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGAGCAGGCTGGGCCCATCTCGGCTATGAGGGCGGAAGTGGGGCGGTTGGGGAGCC 60
DB 1 TGGGAGCAGGCTGGGCCCATCTCGGCTATGAGGGCGGAAGTGGGGCGGTTGGGGAGCC 60

QY 61 TCGTGGCCCTGGCCCATCTCCAGTCCCAGTCTCGGCTCGGACAGATAGGGCGAGGCT 120
DB 61 TCGTGGCCCTGGCCCATCTCCAGTCCCAGTCTCGGCTCGGACAGATAGGGCGAGGCT 120

QY 121 GTGCTGTCTTTTTCAGAAAGACTTCCGGGCGCACCATGCGATCCAGCATCCTCAAGTGGJTG 180
DB 121 GTGCTGTCTTTTTCAGAAAGACTTCCGGGCGCACCATGCGATCCAGCATCCTCAAGTGGGJTG 180

QY 181 TCAGGCCACAGAGCTCGACGAGCAGCAGAGCAAGCCAGGGACACAGAGGAGGAGG 240
DB 181 TCAGGCCACAGAGCTCGACGAGCAGCAGAGCAAGCCAGGGACACAGAGGAGGAGG 240

QY 241 CCGGGAGCAGCAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 300
DB 241 CCGGGAGCAGCAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 300

QY 301 GCCTCCGGGGGAGCTTGTGGACATCGCTGCAAGTGTGCCAGGCTACCTGGGGCAGC 360
DB 301 GCCTCCGGGGGAGCTTGTGGACATCGCTGCAAGTGTGCCAGGCTACCTGGGGCAGC 360

QY 361 TGGAGCATGAGGACATCGACACCTGAGCAGATGCCCGTGAGGACCTCACTGAGGGCGAGT 420
DB 361 TGGAGCATGAGGACATCGACACCTGAGCAGATGCCCGTGAGGACCTCACTGAGGGCGAGT 420

QY 421 GGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATT 480

DB 421 GGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATT 480

QY 481 CAAGAAATTTACTTTTCGCGAGTCCAGGCTCTGCTGAATAGAATCAGCTCTGTGTAACCCCTC 540

DB 481 CAAGAAATTTACTTTTCGCGAGTCCAGGCTCTGCTGAATAGAATCAGCTCTGTGTAACCCCTC 540

QY 541 AGACGGACATTTGACGGGCTCCCGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600

DB 541 AGACGGACATTTGACGGGCTCCCGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600

QY 601 GCCGAGACATAGTGTGCATFGAACCGGTGGAGAGATCTCGAGCTGGGAGCTGCAGACC 660

DB 601 GCCGAGACATAGTGTGCATFGAACCGGTGGAGAGATCTCGAGCTGGGAGCTGCAGACC 660

QY 661 ACCCTCTTTTCCAGGAGCAACAGTGGGTCCAGAGTACATCGAGAGCGCGCTCTCA 720

DB 661 ACCCTCTTTTCCAGGAGCAACAGTGGGTCCAGAGTACATCGAGAGCGCGCTCTCA 720

QY 721 TCTGTGACACCAAGTTCCGACATCAGACAGTGGTTCCTCTCACGAGCTGGAACCCCTGA 780

DB 721 TCTGTGACACCAAGTTCCGACATCAGACAGTGGTTCCTCTCACGAGCTGGAACCCCTGA 780

QY 781 CCATCTGGTCTTACAAGGAGATTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTGGACA 840

DB 781 CCATCTGGTCTTACAAGGAGATTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTGGACA 840

QY 841 AGCTGGACAGCGCCATCCACCTGTGCACAAACGCGTCCAGAGTACCTGGAAGATGATG 900

DB 841 AGCTGGACAGCGCCATCCACCTGTGCACAAACGCGTCCAGAGTACCTGGAAGATGATG 900

QY 901 TGGGGCGCAGCCCTGTCTGCCCGCACACAACATGTGGACACAGCTTCCAGGAGT 960

DB 901 TGGGGCGCAGCCCTGTCTGCCCGCACACAACATGTGGACACAGCTTCCAGGAGT 960

QY 961 ACTTCAGCGCCAGCGGCTGGCGGCTGTGGGAGCGTCACTTACCCGTCATGAAGA 1020

DB 961 ACTTCAGCGCCAGCGGCTGGCGGCTGTGGGAGCGTCACTTACCCGTCATGAAGA 1020

QY 1021 AGGCCATCGCCACCGCATGAAGTGGCCAGGACACCTGGAGCTCCAGAGACAGCT 1080

DB 1021 AGGCCATCGCCACCGCATGAAGTGGCCAGGACACCTGGAGCTCCAGAGACAGCT 1080

QY 1081 TTGAGCTCTACCGGGCTGACTTTCGCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGA 1140

DB 1081 TTGAGCTCTACCGGGCTGACTTTCGCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGA 1140

QY 1141 TCAATTCAGGCCCATCGACCCGTCACCGCGGTCAAGGCCACAGCTGTGTGACAGG 1200

DB 1141 TCAATTCAGGCCCATCGACCCGTCACCGCGGTCAAGGCCACAGCTGTGTGACAGG 1200

QY 1201 TGCAGGAGGACACCATCAAGTGGCGGTGACCGCAGCTGTGACATCGGCAACTTCGAGC 1260

DB 1201 TGCAGGAGGACACCATCAAGTGGCGGTGACCGCAGCTGTGACATCGGCAACTTCGAGC 1260

QY 1261 TCTCTGTGAGGAGCGCGGTGTGAGCCGCCCATTTACAGCGGCTCCGACCTCTCGGTG 1320

DB 1261 TCTCTGTGAGGAGCGCGGTGTGAGCCGCCCATTTACAGCGGCTCCGACCTCTCGGTG 1320

QY 1321 CGGGGCTCAGTGTGAGGAGCGCAGGTGTCGCCGTTCGCAACTTCAGGCTTCAGGCT 1380

DB 1321 CGGGGCTCAGTGTGAGGAGCGCAGGTGTCGCCGTTCGCAACTTCAGGCTTCAGGCT 1380

QY 1381 CGGCTCTCGTGTGAGCGCGCAGCGCTGAAGGACAGGGGCCCTCGGCGCATGCCAGACC 1440

DB 1381 CGGCTCTCGTGTGAGCGCGCAGCGCTGAAGGACAGGGGCCCTCGGCGCATGCCAGACC 1440

QY 1441 CTGCCAGGAGACCCCATCACCAGCTCTCCAGCGGACTTGGGACTTGAAGGAGGAAGG 1500

DB 1441 CTGCCAGGAGACCCCATCACCAGCTCTCCAGCGGACTTGGGACTTGAAGGAGGAAGG 1500

QY 1501 GGCTCCCTCGCTTGTGGCACCTTTAAGGGGGGAGCCGAGGAGCTGGAGCGGCAC 1560

Db	1501	GGCTCCCTCCCTGGCTTGTCTGCACACCTTTAAGGGGGGGCAGCCGAGAGCGGTGGAGCCCGCAC	1561
Qy	1561	AGCCCCACCGCACCAAAAGCTGCTGGGAAGGTGGAGCTCCCGCGCTGCCCCCTGTGTGCGCCACG	1620
Db	1561	AGCCCCACCGCACCAAAAGCTGCTGGGAAGGTGGAGCTCCCGCGCTGCCCCCTGTGTGCGCCACG	1620
Qy	1621	TGGACAGCTCAGGCCCCCAAAACACCGGTGTCCCCGTAGCGCGCAGCCCGCCCAAAGCTGGGATC	1680
Db	1621	TGGACAGCTCAGGCCCCCAAAACACCGGTGTCCCCGTAGCGCGCAGCCCGCCCAAAGCTGGGATC	1680
Qy	1681	CAAAACACAGCTAAATTCGCGCACCGCTGGAGCTGTGCTCGGGGGCTCGAAGACAGCAGCAGG	1740
Db	1681	CAAAACACAGCTAAATTCGCGCACCGCTGGAGCTGTGCTCGGGGGCTCGAAGACAGCAGCAGG	1740
Qy	1741	GC CGCTGCGCTCCGCGCCCGCGAGAAAAGGTTTCATGACACGCTCAGATTCTCTGCGACGA	1800
Db	1741	GC CGCTGCGCTCCGCGCCCGCGAGAAAAGGTTTCATGACACGCTCAGATTCTCTGCGACGA	1800
Qy	1801	GGAGTACAGGTTTGACGACCACTCTCTCCCAAGGGCGAATTTC	1838
Db	1801	GGAGTACAGGTTTGACGACCACTCTCTCCCAAGGGCGAATTTC	1838
RESULT 2			
ADJ93364			
ID ADJ93364 standard; DNA; 3554 BP.			
XX			
AC	ADJ93364;		
XX			
XX			
DT	06-MAY-2004 (first entry)		
XX			
DE	Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.		
XX			
KW	testis-specific tubulin tyrosine-ligase-like polypeptide;		
KW	BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;		
KW	neuroprotective; endocrine-Gen; antinflammatory; anabolic; hypercensive;		
KW	osteopathic; nootropic; antiParkinsonian; antiarthritic; antiasthmatic;		
KW	anti-HIV; antibacterial; immunosuppressive; antiseborrheic;		
KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;		
KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;		
KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;		
KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;		
KW	brain cancer; liver cancer; proliferative condition; testis; lung;		
KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;		
KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;		
KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;		
KW	sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.		
XX			
OS	Homo sapiens.		
OS	Synthetic.		
XX			
PH			
Key	Location/Qualifiers		
FT	CDS		
FT	1695..3320		
FT	/*tag= a		
FT	/product= "Human tubulin tyrosine ligase protein"		
XX			
XX	WO2004005487-A2.		
PN			
XX			
PD	15-JAN-2004.		
XX			
XX	09-JUL-2003; 2003WO-US021605.		
PP			
XX			
PR	09-JUL-2002; 2002US-0394725P.		
XX			
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.		
XX			
PI	Feder JN, Wu S, Nelson TC;		
XX			
XX	WPI: 2004-099381/10.		
DR	P-PSDB; ADJ93365.		
XX			
PT	New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,		
PT	useful for preventing, treating or ameliorating a medical condition, e.g.		

aberrant cellular proliferation, reproductive disorders or testicular disorders.

Example 4; SEQ ID NO 12; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase -like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cytostatic, respiratory -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic, antiapoptotic, antiarthritic, antiasthmatic, anti-HIV, antibacterial, immunosuppressive, antiseborrheic or dermatological activity acting as tyrosine ligase modulators. In addition, the disclosed sequences may be useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of the tubulin tyrosine ligase protein consensus gene sequence which was used in the exemplification of the invention.

Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;

Query Match 97.0%; Score 1782.8; DB 12; Length 3554;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1824; Conservative 0; Mismatches 2; Indels 30; Gaps 1;

QY 1 TGGGAGCAGCCCTGGGCCCCCATCGGCTATCAGGCGCGGAAGTGGGGGGGTGGGGAGCC 60
DB 1513 TGGGAGCAGCCCTGGGCCCCCATCGGCTATCAGGCGCGGAAGTGGGGGGGTGGGGAGCC 1572

QY 61 TCGGCTGGCCCTGGCCCCCATCTCCAGTCCCAGTCTCGGCTCGGACAGATAGGCGAGGCT 120
DB 1573 TCGGCTGGCCCTGGCCCCCATCTCCAGTCCCAGTCTCGGCTCGGACAGATAGGCGAGGCT 1632

QY 121 GTGCTGTCTTTTCA-----GAAAGCTTCGGCGCA 150
DB 1633 GTGCTGTCTTTTTCAGAGTAGCAGCGTGGGAAGGAGTTCCTCGGAAGACTTCGGCGCA 1692

QY 151 CCATGCATCCAGTCCTCAAGTGGGTGTGAGCCACAGAGCTGCAGCAGAGGAGCA 210
DB 1693 CCATGGCATCCAGTCCTCAAGTGGGTGTGAGCCACAGAGCTGCAGCAGAGGAGCA 1752

QY 211 GAAGCAAGCCCGAGCACAGAGGAGGAGCCCGGAGCAGGACCTGAGCAGCAGGCAAG 270
DB 1753 GAAGCAAGCCCGAGGACCAAGAGGAGGAGCCCGGAGCAGGACCTGAGCAGCAGGCAAG 1812

QY 271 ATGCTGAAAATGTGAGGCAAGACTCAGGGGCTCCCGGGGAGCTTTGTGGAATCGCGT 330
DB 1813 ATGCTGAAAATGTGAGGCAAGACTCAGGGGCTCCCGGGGAGCTTTGTGGAATCGCGT 1872

QY 331 GCAAGGTGCGCAGGCTTACCTGGGCGAGCTGAGCATGAGGACATCGACACGTCAGCAG 390
DB 1873 GCAAGGTGCGCAGGCTTACCTGGGCGAGCTGAGCATGAGGACATCGACACGTCAGCAG 1932

QY 391 ATCCGCTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACCCAGCAGTACTACTCCC 450
DB 1933 ATCCGCTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACCCAGCAGTACTACTCCC 1992

QY 451 TCGTTTCATGCGCATGCTTTTCATCTCCAATTCAGAAATTACTTTTCGAGTGCAGGCTC 510

Db 1993 TCCTTCATGCGGATGCTTTTCATCTCCAAATTCAGAAATTAATTTTCGACAGTGCCAGGCTC 2052
QY 511 TGTCTGAATGAATACAGCTCTGTGAACCCCTCAGACGACATTTAGACGGGCTCCGGAACATCT 570
Db 2053 TGTCTGAATGAATACAGCTCTGTGAACCCCTCAGACGACATTTAGACGGGCTCCGGAACATCT 2112
QY 571 GGATTATAAGCCCCGGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGGTGTG 630
Db 2113 GGATTATAAGCCCCGGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGGTGTG 2172
QY 631 AGGAGATCTTGGAGCTGGAGCTGGAGACCAACCTCTTCCAGGGACAAAGTGGGTGG 690
Db 2173 AGGAGATCTTGGAGCTGGAGCTGGAGACCAACCTCTTCCAGGGACAAAGTGGGTGG 2232
QY 691 TCCAGAAGTACATCGAGACGCCCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 750
Db 2233 TCCAGAAGTACATCGAGACGCCCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 2292
QY 751 GGTTCCTCGTCAAGGACTGGAACCCCTGACCAATCTGGTTTCTAAGAGGAGTTACTTGC 810
Db 2293 GGTTCCTCGTCAAGGACTGGAACCCCTGACCAATCTGGTTTCTAAGAGGAGTTACTTGC 2352
QY 811 GGTTCCTCAACTAGCGCTTCTCCTCGACNAGCTGGACAGCGCCATCCACCTGTGCAACA 870
Db 2353 GGTTCCTCAACTAGCGCTTCTCCTCGACNAGCTGGACAGCGCCATCCACCTGTGCAACA 2412
QY 871 ACGCCGTCCAGAAAGTACCTGAAGAAATGATGTGGCCGCGACAGCCCTGCTGCCGACACA 930
Db 2413 ACGCCGTCCAGAAAGTACCTGAAGAAATGATGTGGCCGCGACAGCCCTGCTGCCGACACA 2472
QY 931 ACATGTGGACACGACACAGGTTTCCAGAGGTACCTGCAGCGCCGAGGGCCGTGGCGCCGTGT 990
Db 2473 ACATGTGGACACGACACAGGTTTCCAGAGGTACCTGCAGCGCCGAGGGCCGTGGCGCCGTGT 2532
QY 991 GGGGACGCTCATCTTACCGTTCATGAAGGCCATGGCCACGCGCATGAAGTGGCC 1050
Db 2533 GGGGACGCTCATCTTACCGTTCATGAAGGCCATGGCCACGCGCATGAAGTGGCC 2592
QY 1051 AGGACACGTGGAGCTTCGCAAGAACAGCTTTTCAGCTCTACGGGGCTGACTTCGTCTTGTG 1110
Db 2593 AGGACACGTGGAGCTTCGCAAGAACAGCTTTTCAGCTCTACGGGGCTGACTTCGTCTTGTG 2652
QY 1111 GGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCCCCACCATCGACCCCGTCCA 1170
Db 2653 GGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCCCCACCATCGACCCCGTCCA 2712
QY 1171 CGCCGCTCAGGCCCGCAGCTGTGTGCACAGTGCAGGAGACACCATCAAGTGGCCGTGG 1230
Db 2713 CGCCGCTCAGGCCCGCAGCTGTGTGCACAGTGCAGGAGACACCATCAAGTGGCCGTGG 2772
QY 1231 ACGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGACGCGGTGGTTGAGCCCG 1290
Db 2773 ACGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGACGCGGTGGTTGAGCCCG 2832
QY 1291 CCCCAITTCAGCGGGTCCGACCTCTGCGTGGCGGGCGTCAAGTGTGAGGAGAGCCAGGAGC 1350
Db 2833 CCCCAITTCAGCGGGTCCGACCTCTGCGTGGCGGGCGTCAAGTGTGAGGAGAGCCAGGAGC 2892
QY 1351 AGGTGTCGCCCTGTGCAACCTCAAGGCTCGGCCCTCGCTTGTGGACGCGCAGCCCTGA 1410
Db 2893 AGGTGTCGCCCTGTGCAACCTCAAGGCTCGGCCCTCGCTTGTGGACGCGCAGCCCTGA 2952
QY 1411 AGGCAGCGGGCCCTCGGCCATGCCAGACCCCTGCCAGGGACCCCATCACAGCTCTCC 1470
Db 2953 AGGCAGCGGGCCCTCGGCCATGCCAGACCCCTGCCAGGGACCCCATCACAGCTCTCC 3012
QY 1471 AGCGGACTTGGGACTGAAGAGAGAGAGGGGCTCCCTCGCTTGTGCTGGCACCCCTTAA 1530
Db 3013 AGCGGACTTGGGACTGAAGAGAGAGAGGGGCTCCCTCGCTTGTGCTGGCACCCCTTAA 3072
QY 1531 GGGGGGACCGGAGAGCGGTGGAGCGGCGACAGCCACCGCACAAGTGTGGGAAGG 1590
Db 3073 GGGGGGACCGGAGAGCGGTGGAGCGGCGCACAGCCACCGCACAAGTGTGGGAAGG 3132

QY 1591 TGGAGCTCCCGGCTCCCTCTGTCGCCACGTGGACAGTCAGGCCCAAAACACCGGTGTCC 1650
Db 3133 TGGAGCTCCCGGCTCCCTCTGTCGCCACGTGGACAGTCAGGCCCAAAACACCGGTGTCC 3192
QY 1651 CCGTAGCCACGAGCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGCGGACCCGCTGGAGC 1710
Db 3193 CCGTAGCCACGAGCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGAGCACCCGCTGGAGC 3252
QY 1711 CTGTGCTGGGGCCCTGAAGACAGCAGAGGGCGCTGCTGCCGCCGCCGAGGAAAAG 1770
Db 3253 CTGTGCTGGGGCCCTGAAGACAGCAGAGGGCGCTGCTGCCGCCGCCGAGGAAAAG 3312
QY 1771 GTTCATGACAGCGCTCAGATTCTCTGCAGCAGGAGTACAGGTTGCAGCCACTCTCCC 1826
Db 3313 GTTCATGACAGCGCTCAGATTCTCTGCAGCAGGAGTACAGGTTGCAGCCACTCTCCC 3368
RESULT 3
ADJ93363
ID ADJ93363 standard; DNA; 3465 BP.
XX
AC ADJ93363;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone C SeqID11.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; stomach cancer; lung cancer;
gastrointestinal disorder; colon cancer; pulmonary cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; db.
XX
OS Homo sapiens.
XX
PN WO2004005487-A2.
XX
PD 15-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
DR WPI; 2004-099381/10.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
useful for preventing, treating or ameliorating a medical condition, e.g.
aberrant cellular proliferation, reproductive disorders or testicular
disorders.
XX
PS Example 4; SEQ ID NO 11; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
-like polypeptide, designated the BGS-42 polypeptide. The invention may
be useful for the development of compounds with a cytostatic, respiratory
-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC
antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC
immunosuppressive, antiseborrheic or dermatological activity acting as

CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.

XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match 91.3%; Score 1678.8; DB 12; Length 3465;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1757; Conservative 0; Mismatches 2; Indels 67; Gaps 1;

QY 1 TGGAGCAGCGCTGGGCCCCATCGGCTATCAGGGCGGGAAGTGGGGGGTTGGGGAGCC 60
DB 1526 TGGAGCAGCGCTGGGCCCCATCGGCTATCAGGGCGGGAAGTGGGGGGTTGGGGAGCC 1585
QY 61 TCGGTGCGCCCTGGCCCCATCTCCAGTCCCGAGTCCCGGCTCGGACAGATAGGCGGAGGCT 120
DB 1586 TCGGTGCGCCCTGGCCCCATCTCCAGTCCCGAGTCCCGGCTCGGACAGATAGGCGGAGGCT 1645
QY 121 GTGCTGTCTTTTTCAGAGACTTCGGCGCACCATGGCATCCAGCATCTCTCAAGTGGGTGG 180
DB 1646 GTGCTGTCTTTTTCAGAGACTTCGGCGCACCATGGCATCCAGCATCTCTCAAGTGGGTGG 1705
QY 181 TCAGCCACAGAGCTGACAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 1706 TCAGCCACAGAGCTGACAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1765
QY 241 CCGGGAGCAGCGACCTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 300
DB 1766 CCGGGAGCAGCGACCTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1825
QY 301 GCCTCCCGGGGCGAGCTTGTGGACATCGCGTGCAGAGTGTGCCAGGCGCTACTTGGGGCAGC 360
DB 1826 GCCTCCCGGGGCGAGCTTGTGGACATCGCGTGCAGAGTGTGCCAGGCGCTACTTGGGGCAGC 1885
QY 361 TGGAGCATAGGACATCGACACGTGACAGATGCGCGTGGAGAGCCTCAGTGGGCGGAGT 420
DB 1886 TGGAGCATAGGACATCGACACGTGACAGATGCGCGTGGAGAGCCTCAGTGGGCGGAGT 1945
QY 421 GGGAGGACCTGACCCAGCAGTACTTCTCCCTGCTCATGGCGATGCTTTCATCTCCAAAT 480
DB 1946 GGGAGGACCTGACCCAGCAGTACTTCTCCCTGCTCATGGCGATGCTTTCATCTCCAAAT 2005
QY 481 CAAGAAATTAATTTTTCGACGTGCGAGGCTCTGCTGAATAGAATCAGCTCTGTGAACCTTC 540
DB 2006 CAAGAAATTAATTTTTCGACGTGCGAGGCTCTGCTGAATAGAATCAGCTCTGTGAACCTTC 2065
QY 541 AGACGACATGACGGGCTCCGAAACATCTGGATTATAAGCCCGGCGCAAGTCCCGGG 600
DB 2066 AGACGACATGACGGGCTCCGAAACATCTGGATTATAAGCCCGGCGCAAGTCCCGGG 2125
QY 601 GCCGAGACATAGTGTGCATGGAACCGTGTGGAGAGATCTGGAGCTGGGAGCTGCAGACC 660
DB 2126 GCCG----- 2129
QY 661 ACCCTCTTTCAGGGACAAAGTGGGTGGTCCAGAAATACATCGAGAGCGCCGCTGCTCA 720
DB 2130 -----AGGGACACAAAGTGGGTGGTCCAGAAATACATCGAGAGCGCCGCTGCTCA 2178

QY 1801 GGAGTACAGTTTCAGCCACTCTCCC 1826
DB |||||
DB 3259 GGAGTACAGTTTCAGCCACTCTCCC 3284

RESULT 4

ID ADJ93361
AC ADJ93361 standard; DNA; 1939 BP.

XX ADJ93361;

DT 06-MAY-2004 (first entry)

XX Human BGS-42 protein-related DNA clone A SeqID9.

DE testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.

XX Homo sapiens.

OS WO2004005487-A2.

PN 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

PF 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.

PS Example 4; SEQ ID NO 9; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytosolic, respiratory
XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
XX antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
XX antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
XX neural disorders, brain cancer, liver cancer, or proliferative condition
XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
XX polypeptide, polynucleotide, or their modulators are also useful for
XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's

CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.

XX SQ Sequence 1939 BP; 421 A; 592 C; 589 G; 337 T; 0 U; 0 Other;

Query Match 91.3%; Score 1677.2; DB 12; Length 1939;

Best Local Similarity 96.2%; Pred. No. 0;

Matches 1756; Conservative 0; Mismatches 3; Indels 67; Gaps 1;

QY 1 TGGGAGCAGGCTGGGCCCCCATCGGCTATAGCGGGGAAAGTGGGGCGGTTTGGGAGCC 60
DB |||||

QY 61 TCGTGGCCCTGGCCCCCATCTCCAGTCCCAGTCTCGTCGGTCGGACAGATAGGCGGAGGCT 120
DB |||||

QY 61 TCGTGGCCCTGGCCCCCATCTCCAGTCCCAGTCTCGTCGGTCGGACAGATAGGCGGAGGCT 120
DB |||||

QY 121 GTGCTGTCTCTTTCAGAGACTTTCCGGCGCACCATGGCATCCAGCATCTCTCAAGTGGTGG 180
DB |||||

QY 121 GTGCTGTCTCTTTCAGAGACTTTCCGGCGCACCATGGCATCCAGCATCTCTCAAGTGGTGG 180
DB |||||

QY 181 TCAGCCACCAGAGCTGCGAGGAGCAGCAGAAGCAAGCCAGGAGCCAGAGGGAGG 240
DB |||||

QY 181 TCAGCCACCAGAGCTGCGAGGAGCAGCAGAAGCAAGCCAGGAGCCAGAGGGAGG 240
DB |||||

QY 241 CCGGAGCAGCGACCTGAGCAGCAGCAAGATGCTGAAAATCTGAGGCAAGACTCAGGG 300
DB |||||

QY 241 CCGGAGCAGCGACCTGAGCAGCAGCAAGATGCTGAAAATCTGAGGCAAGACTCAGGG 300
DB |||||

QY 301 GCCTCCCGGGCAGCTTGTGACATCGCGTGAAGGTGTCCAGGCTACCTGGGGCAGC 360
DB |||||

QY 301 GCCTCCCGGGCAGCTTGTGACATCGCGTGAAGGTGTCCAGGCTACCTGGGGCAGC 360
DB |||||

QY 361 TGGAGCATGAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGT 420
DB |||||

QY 361 TGGAGCATGAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGT 420
DB |||||

QY 421 GGGAGGACCTGACCCAGCAGTACTCTCCCTGTTTCATGGCGATGCTTCTCAATTC 480
DB |||||

QY 421 GGGAGGACCTGACCCAGCAGTACTCTCCCTGTTTCATGGCGATGCTTCTCAATTC 480
DB |||||

QY 481 CAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCAGCTGTGAACCTTC 540
DB |||||

QY 481 CAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCAGCTGTGAACCTTC 540
DB |||||

QY 541 AGACGGACATTGACGGGCTCCGGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600
DB |||||

QY 541 AGACGGACATTGACGGGCTCCGGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600
DB |||||

QY 601 GCGGAGACATAGTGTGATGGACCTGTGGAGGAGATCCTGGAGCTGGCAGTGCAGACC 660
DB |||||

QY 601 GCGGAGACATAGTGTGATGGACCTGTGGAGGAGATCCTGGAGCTGGCAGTGCAGACC 660
DB |||||

QY 661 ACCCTCTTTTCCAGGAGCAACAGTGGGTGCTCAGAAATACATCGAGAGCCGCTGCTCA 720
DB |||||

QY 605 -----AGGACAAACAGTGGGTGCTCAGAAATACATCGAGAGCCGCTGCTCA 653
DB |||||

QY 721 TCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTACGGAATGGAACCCCTGA 780
DB |||||

QY 654 TCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTACGGAATGGAACCCCTGA 713
DB |||||

QY 781 CCATCTGTTTTCAGAGGAGTACTTCTCACTTCAGCGCTTCTCCCTGGACA 840
DB |||||

QY 714 CCATCTGTTTTCAGAGGAGTACTTCTCACTTCAGCGCTTCTCCCTGGACA 773
DB |||||

QY 841 AGCTGGACAGCGCCATCCACCTGTGCAACCAACCCGTCAGAAAGTACCTGAAGAAATGATG 900
DB |||||

QY 774 AGCTGGACAGCGCCATCCACCTGTGCAACCAACCCGTCAGAAAGTACCTGAAGAAATGATG 833
DB |||||

PS Example 2; SEQ ID NO 631; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the

CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.

CC The current sequence is that of a human soft tissue sarcoma-upregulated

CC DNA of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.

XX

SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;

Query Match 27.3%; Score 501; DB 12; Length 101270;

Best Local Similarity 100.0%; Pred. No. 6.3e-97;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1272 CAGCCGGTGTTCAGCGCGCCCATTCAGCGGTCCGACTCTGCGTGGGGCGTCAGT 1331

DB 43637 CAGCCGGTGTTCAGCGCGCCCATTCAGCGGTCCGACTCTGCGTGGGGCGTCAGT 43578

QY 1332 GTGAGGAGACCCAGGAGGAGTGTGCTGCCGCTCTGCAACTCAAGGCTCGGCTCGCTG 1391

DB 43577 GTGAGGAGACCCAGGAGGAGTGTGCTGCCGCTCTGCAACTCAAGGCTCGGCTCGCTG 43518

QY 1392 TTGGACGCGCAGCGCTGAAGGCAACGGGCGCCCTCGGGCCATGCGACCTGCCCCAGGGA 1451

DB 43517 TTGGACGCGCAGCGCTGAAGGCAACGGGCGCCCTCGGGCCATGCGACCTGCCCCAGGGA 43458

QY 1452 CCCCCATCACCAGCTCTCCAGCGGGACTTGGGACTCAAGAGAGAGGGGCTCCCCCTG 1511

DB 43457 CCCCCATCACCAGCTCTCCAGCGGGACTTGGGACTCAAGAGAGAGGGGCTCCCCCTG 43398

QY 1512 GCCTTGTGTCACCCCTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCG 1571

DB 43397 GCCTTGTGTCACCCCTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCG 43338

QY 1572 ACCAAGAGCTGTGGGAAGTGTGAGCTCCGCGCTCGGCTGCTGCCACGTGGACATCAG 1631

DB 43337 ACCAAGAGCTGTGGGAAGTGTGAGCTCCGCGCTCGGCTGCTGCCACGTGGACATCAG 43278

QY 1632 GCCCCCAACACCCGCTGCCCGTAGCCGCGCCGACCAAGCTGGGATCCCAACACGCTA 1691

DB 43277 GCCCCCAACACCCGCTGCCCGTAGCCGCGCCGACCAAGCTGGGATCCCAACACGCTA 43218

QY 1692 AATGCGCACCCGCTGGAGCTGTGCTGCGGGGCTTGAAGACAGCAGAGGGCGGCTGCCGT 1751

DB 43217 AATGCGCACCCGCTGGAGCTGTGCTGCGGGGCTTGAAGACAGCAGAGGGCGGCTGCCGT 43158

QY 1752 CCGCCGCCCGGAGGAAAAGT 1772

DB 43157 CCGCCGCCCGGAGGAAAAGT 43137

RESULT 8

AAS99894

ID AAS99894 standard; cDNA; 2380 BP.

XX

AC AAS99894;

XX

DT 12-MAR-2002 (first entry)

XX

DE Polynucleotide encoding human cytoskeleton-associated protein (CYSKP) #5.

XX

KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;

KW cell proliferative disorder; inflammatory disorder; prion disease;

KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;

KW neurological disorder; cell motility disorder; reproductive disorder;

KW spinal cord disease; central nervous system disorder; mental disorder;

KW gene therapy; cancer.

XX

OS Homo sapiens.

XX WO200185942-A2.

XX

PD 15-NOV-2001.

XX

PF 03-MAY-2001; 2001WO-US014355.

XX

PR 05-MAY-2000; 2000US-0201960P.

PR 08-MAY-2000; 2000US-0202729P.

PR 05-JUN-2000; 2000US-0209705P.

PR 07-JUN-2000; 2000US-0210149P.

PR 21-JUN-2000; 2000US-0213215P.

XX

XX (INCY-) INCYTE GENOMICS INC.

XX

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;

PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;

PI Policky JL;

XX

DR WPI; 2002-062248/08.

DR P-PSDB; AAU74334.

XX

XX New cytoskeleton-associated proteins and polynucleotides, useful for

PT diagnosing, preventing and treating cell proliferative, autoimmune,

PT inflammatory, neurological, cell motility, reproductive and muscle

PT disorders.

XX

PS Claim 5; Page 171; 194pp; English.

XX

CC The invention relates to human cytoskeleton-associated polypeptides

CC (CYSKP) and their associated polynucleotide sequences. The sequences are

CC useful in the treatment of disorders associated with overexpression or

CC underexpression of CYSKP in a patient. The disorders include cell

CC proliferative disorders (such as cancer, actinic keratosis,

CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),

CC autoimmune/inflammatory disorders (such as asthma, atherosclerosis,

CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus

CC and anaemia), vesicle trafficking disorders (such as

CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),

CC gastrointestinal disorders, prion diseases, neurological disorders (such

CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,

CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis

CC and other motor neuron disorders), cell motility disorders, reproductive

CC disorders (such as endometriosis and polycystic ovary syndrome), muscle

CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,

CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord

CC diseases, central nervous system disorders (such as Down syndrome and

CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

CC Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP

CC of the invention

XX

SQ Sequence 2380 BP; 580 A; 685 C; 670 G; 445 T; 0 U; 0 Other;

Query Match 23.5%; Score 432.8; DB 6; Length 2380;

Best Local Similarity 59.1%; Pred. No. 1.1e-82;

Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;

QY 260 CAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCCTCCGGGGCAGCTTG 318

DB 361 CAGGAGACAAGCAGCCCAAGAACAGGAGAAAAACCCAGTGTGGTGTCCCCAGAGTTTG 420

QY 319 TGGACATCGCGTGAAGGTGTCCAGGCTACTCTGGGCGAGCTGGAGCATGAGCATCG 378

DB 421 TGGATGAAGCTCTGTGTGCGTGGAGGAGTACCTTAGCAACTTGGCCCATGAGCATCG 480

QY 379 ACACGCTCAGCAGATGCGGTGGAGGACCTCAGTCAGGCCGAGTGGGAGGACCTGACCCAGC 438

DB 481 ACAAGACCTGGAGGCGCCGCTGTACTCACCCCGAGGGCTGGTCCCTCTTCTCCACGC 540

Query Match		23.5%; Score 431.6; DB 4; Length 2326;
Best Local Similarity		63.5%; Pred. No. 2e-82;
Matches		698; Conservative 0; Mismatches 389; Indels 13; Gaps 2;
QY	260	CAGCAGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCTCCGGGGCAGCTTG 318
DB	620	CAGAGACAAGCAGCCCAAGAAACAGAGAAAACCCAGTGTGGTGTCCCGAGAGTTTG 679
QY	319	TGGACATCGCGTGCAAGGTGTGCAGGCCCTTACCTGGGGCAGCTGGAGCATGAGGACATCG 378
DB	680	TGGATGAAGCTCTGTGCTGTGCGAGGAGTACCTTAGCAACTTGGCCCATGACATCG 739
QY	379	ACAGCTCAGCAGATCGCGTGGAGGACCTCTACTAGGCCGAGTGGAGGACCTGACCCAGC 438
DB	740	ACAAGGACCTTGGAGGGCCCGCTGTACTACCCCGGAGGGCTGTCTCTCCAGC 799
QY	439	AGTACTACTCCCTCGTTCTAGGGGATGCTTTCATCTCCRAATTCAGAAATTACTTTTCG 498
DB	800	GCTACTACCAAGTGTCTCAGAAAGGGCGAGAACTCAGGCACTCTGACACTCAGGTCAGC 859
QY	499	AGTGCCAGGCTCTGCTGAATAGAAATCACGCTCTGTGAACCCCTCAGACGGACATTTGACGGG 558
DB	860	GCTGTGAGGACATCTCTGCAGCAGCTGCAGGCCGTGGTACCCAGATAGACATGGAAGGG 919
QY	559	TCCGGAACTCTCGATTATAAAGCCCGCGGCCAAGTCCCGGGGCCGAGACATAGTGTGCA 618
DB	920	ATCGCAACATCTGGATCGTAGCCAGGAGCCAAAGTCCCGCGAGCAGGACATCATGTGCA 979
QY	619	TGACCGGTGTGGAGGAGATCTCGAGCTGCGACCTGCAGACCAACCCCTCTTCCAGGGACA 678
DB	980	TGGACCCTCTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGTGAAGAGCG 1039
QY	679	ACAAGTGGGTGTCCAGAAGTATCATCGAGACGCGCTGTCTCATCTGTGACACCAAGTTTCG 738
DB	1040	GCAAGTGGGTGTGCAGAAATATATTAGCGGCCCTCTCTCATCTTTGGCACCAGTTTG 1099
QY	739	ACATCAGACATGTTGCTCTCGTCAAGGACTGGAACCCCTGACCATCTGTTCTTACAGG 798
DB	1100	ACCTCAGACATGTTTCTCGTAACTGACTGGAACCCACTTACCGTGTGTTCTACCGCG 1159
QY	799	AGAGTTACTTGGCGTTCTCAACTCAGCGTTCTCCCTGGACAGCTGGACGGCCATCC 858
DB	1160	ACAGCTATATCCCGTTTTCACGACGCGCTTCTCCCTGAAGAACCTTGGACAACTCAGTGC 1219
QY	859	ACCTGTGCAACACGCGCTCCAGAAGTACTCTGAAGAAATGATGTGGCGCAGCCCTCTGC 918
DB	1220	ACCTGTGCAACACTCCATCCAGAACACTCTGAGAACTCATGCCATCGGCATCCACTGC 1279
QY	919	TGCCCGCACACAATGTGGACACGACACAGTTTCAGAGGTACCTGCAAGCGCCAGGGCC 978
DB	1280	TTCCGCCAGACAACTGTGGTCTTAGCCAGAGGTTCCAGGCCCACTTGCAGAGATGGGTG 1339
QY	979	GTGGCGCGGTGTGGGCAGCGTCTATCCCGTCCATGAGAGGCCATCGCCACGCCA 1038
DB	1340	CCCCAAATGTTGTTCACCATCATCTGCTGGCATGAAGGATGCTGTGATCCAGCAC 1399
QY	1039	TGAAGTGGGCCAGGACCACTGGAGCCTCGCAAGAACAGCTTTTGAAGCTCTACGGGGCTG 1098
DB	1400	TTCAGACCTCCAGGACACCGTCACTGTCGAGAGCCAGCTTTTGAAGCTCTATGGCGCTG 1459
QY	1099	ACTTCGTCTTTGGAGGGACTTCAGGCCCTTGGCTGTGATCGAGATCAATTCAGGCCCA 1158
DB	1460	ACTTCGTGTTCGGGGAGGACTTCAGGCCCTTGGCTGTGATCGAGATCAACGGCCAGCC 1519
QY	1159	TGCACCGCTCCAGCGGTCAAGGCTGCTGTCAGAGTGCACAGGTGAGGAGGACCATCA 1218
DB	1520	TGGACCCCTCCAGCAGATCACTGCGCCGGCTGTGTCTGGCGTGAAGTGAACCCCTGC 1579
QY	1219	AGGTGGCC-----GTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGT 1266
DB	1580	GCGTGTCTTACCGGATCTGACCGCAACTGTGACACAGGAGCTTTGAGCTCATCT 1639
QY	1267	GGAGGACAGCGGTGTTGAGCGCCGCCCAATTCAGCGGGTCCGACCTCTCGTGGCGGGCG 1326

DB	1640	ATAAGCAGCTCTCTGGAGTGCCTCAATATGTGGGCATCGCGCTCTCTGGTAGAGGCT 1699
QY	1327	TCAGTGTGAGGAGCCAGG 1346
DB	1700	TCACCATCAAGAGCCCATG 1719
RESULT 10		
ACH91699/c		
ID	ACH91699	standard; DNA; 490 BP.
XX	ACH91699;	
DT	29-JUL-2004	(first entry)
XX	Human genome derived single exon probe #24894.	
XX	Human; probe; ss; gene expression; single exon probe; microarray;	
KW	alternative splicing event; genomic alteration.	
XX	Homo sapiens.	
XX	US2003194704-A1.	
XX	16-OCT-2003.	
XX	03-APR-2002; 2002US-00029386.	
XX	03-APR-2002; 2002US-00029386.	
XX	(PENN/) PENN S G.	
PA	(RANK/) RANK D R.	
PA	(HANZ/) HANZEL D K.	
PI	Penn SG, Rank DR, Hanzel DK;	
XX	WPI; 2004-119264/12.	
XX	New human genome-derived single exon nucleic acid probes useful for human	
PT	gene expression analysis, for identifying or characterizing alternative	
PT	splicing events, for assessing genomic alterations or as tools for	
PT	surveying tissues.	
XX	Claim 1; SEQ ID NO 24894; 80pp; English.	
PS	The invention relates to a nucleic acid probe for measuring human gene	
CC	expression, comprising any of the 27,400 fully defined nucleotide	
CC	sequences in the specification, or their complements or fragments, and	
CC	encoding at least 8 amino acids of any of the 688 amino acid sequences	
CC	fully defined in the specification. The probe is a single exon probe that	
CC	hybridises under high stringency conditions to a nucleic acid molecule	
CC	expressed in human cells or tissues. Also included are a spatially-	
CC	addressable set of single exon nucleic acid probes for measuring human	
CC	gene expression (comprising a plurality of single exon nucleic acid	
CC	probes cited above, where each of the plurality of probes is separately	
CC	and addressably isolatable or amplifiable from the plurality), a single	
CC	exon microarray for measuring human gene expression, a method of	
CC	measuring human gene expression, a vector comprising the single exon	
CC	probe cited above, an ORF-encoded peptide comprising at least 8	
CC	contiguous amino acids of any of the above-mentioned amino acid	
CC	sequences (optionally with conservative amino acid substitutions), an	
CC	isolated antibody that binds specifically to a peptide cited above,	
CC	methods of selling and/or licensing single exon probes or microarrays to	
CC	a customer desiring to measure gene expression, a method of providing	
CC	human gene expression data by subscription, and a computer-readable	
CC	storage medium which contains a database having a plurality of records	
CC	(each record including data on the expression of a single exon probe	
CC	cited above. The probe, methods and apparatus are useful in gene	
CC	expression analysis. The probes may be used as tools for surveying	
CC	tissues to detect the presence of expressed messages that contain their	
CC	specific exon, or in constructing genome-derived single exon microarrays.	
CC	In addition, the probes are used in identifying and characterising	

CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX

XX Sequence 490 BP; 77 A; 150 C; 168 G; 95 T; 0 U; 0 Other;

Query Match 23.2%; Score 427; DB 12; Length 490;

Best Local Similarity 100.0%; Pred. No. 1.3e-81;

Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 848 CAGGGCCATCCTGTGCAACACGCGCTCCAGAGTACTGGAAGTATGTTGGCCG 907

DB 427 CAGGGCCATCCTGTGCAACACGCGCTCCAGAGTACTGGAAGTATGTTGGCCG 368

QY 908 CAGCCCCCTGCTGCCGCACACACATGTGGACACGACAGTTCAGGAGTACCTGCA 967

DB 367 CAGCCCCCTGCTGCCGCACACACATGTGGACACGACAGTTCAGGAGTACCTGCA 308

QY 968 CGCCACGCGCGTGGCGCCGTGTGGGCGACGCTCATCTACCCGTCCATGAAGAGCCAT 1027

DB 307 CGCCACGCGCGTGGCGCCGTGTGGGCGACGCTCATCTACCCGTCCATGAAGAGCCAT 248

QY 1028 CGCCACGCGCATGAAGTGTGCCCGCAGGACACGCTGGAGCTTCGCAAGACGTTTGAGCT 1087

DB 247 CGCCACGCGCATGAAGTGTGCCCGCAGGACACGCTGGAGCTTCGCAAGACGTTTGAGCT 188

QY 1088 CTACGGGGTCACTTCGTCCTGGGAGGACTTCAGGCGCTGCTGATCAGATCAATTC 1147

DB 187 CTACGGGGTCACTTCGTCCTGGGAGGACTTCAGGCGCTGCTGATCAGATCAATTC 128

QY 1148 CAGCCCCCACCATGACCCCGTCCACCGCGTTCAGGCGCCAGCTGTGTGCAGGTGAGGA 1207

DB 127 CAGCCCCCACCATGACCCCGTCCACCGCGTTCAGGCGCCAGCTGTGTGCAGGTGAGGA 68

QY 1208 GGAACACATCAAGGTGGCGCTGTGACCGCAGCTGTGACATCGGCAACTTCGAGTCTCTGTG 1267

DB 67 GGAACACATCAAGGTGGCGCTGTGACCGCAGCTGTGACATCGGCAACTTCGAGTCTCTGTG 8

QY 1268 GAGGCAG 1274

DB 7 GAGGCAG 1

RESULT 11

AD873168

ID AD873168 standard; cDNA; 2538 BP.

XX

AC

XX

DT

XX

XX

DE

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Algate PA, Mannion J, Gaiger A, Gordon B, Harlocker SL;
XX WPI; 2002-759855/82.
XX

XX New isolated polynucleotides and polypeptides, useful for detecting the
XX presence of, and treating cancer, particularly kidney cancer by
XX stimulating T-cells specific for a tumor protein, and stimulating immune
XX response in a patient.

XX Claim 1; SEQ ID NO 1765; 78pp; English.

XX The invention relates to a new isolated polynucleotide (a Human kidney

XX tumour specific cDNA) comprising any one of the 1855 sequences identified

XX in the specification (or their complements, degenerate variants,

XX sequences consisting of at least 20 contiguous residues them, sequences

XX that hybridise to them under highly stringent conditions or sequences

XX having at least 75 or 90% sequence identity to the 1855 sequences. Also

XX included are detecting/determining the presence of cancer in a patient,

XX stimulating an immune response in a patient; treating kidney cancer in a

XX patient, an isolated polypeptide encoded by one of the 1855 sequences, an

XX expression vector comprising the polynucleotide operably linked to an

XX expression control sequence, a host cell transformed/transfected with the

XX vector, an isolated antibody (or its antigen-binding fragment) that

XX specifically binds to the protein, a fusion protein comprising at least

XX one the proteins, stimulating and/or expanding T-cells specific for a

XX tumour protein, an isolated T-cell population comprising the T-cells, a

XX composition comprising a first component (such as a carrier or

XX immunostimulant) and a second component (comprising one of the

XX polynucleotides, the polypeptides, an antibody, T-cell or an antigen-

XX presenting cell that expresses the polynucleotide) and a diagnostic kit

XX comprising at least one of the oligonucleotides, or at least one antibody

XX and a detection reagent comprising a reporter group. The polynucleotides,

XX polypeptides, antibodies and antigen-presenting cells are useful for

XX detecting the presence of, and treating cancer, particularly kidney

XX cancer by stimulating and/or expanding T-cells specific for a tumour

XX protein, and stimulating immune response in a patient. The present

XX sequence is one of the Human kidney tumour specific cDNAs. Note: The

XX sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from USPTO

XX at seqdata.uspto.gov/sequence.html?DocID=20030109434.

XX

XX Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;

Query Match 22.9%; Score 421.4; DB 7; Length 2538;

Best Local Similarity 63.8%; Pred. No. 3e-80;

Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;

QY 232 GGGAGGAGCGCGGAGCAGGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA 291

DB 754 GGGATCGTGACAGGCGCCCTCCCTTATGTGAGGAGACAGAGCCCAAGAACAGGAGAAA 813

QY 292 AGCTCAG-GGGGCTCCCGGGGCGAGCTTGTGGACATCGCGTGCAGAGTGTGCCAGGCGCTAC 350

DB 814 AACCCAGTGTGGTGTCCCGCAGAGTTTGTGATGAAGCTCTGTGTGGTGGAGAGTAC 873

QY 351 CTGGGGCAGCTGAGCATGAGGACATCGACACGTCAGCAGATGCCCTGGAGGACCTCACT 410

DB 874 CTTAGCAACTTGGCCCAATGACATCGACAGGACCTGGAGGCGCCGCTGTACCTCACC 933

QY 411 GAGGCGAGTGGGAGGACCTGACCCAGCAGTACTACTCCTCGTTCATGCGGATGCTTTC 470

DB 934 CCGAGGGCTGTGTCCTCTTCTCCAGCGCTACTACAAAGTGGTGCACGAAAGGGGAGAA 993

QY 471 ATCTCCAATTCAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAATCAAGTCT 530

DB 994 CTCAGGACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTGCGAGGAGTGCAGGCC 1053

QY 531 GTGAACCTTCAGACGACATTTGACGGGCTCCGGGAAACATCTGGATTATAAAGCCCGGGCC 590

DB 1054 GTGGTACCCAGATAGACATGGAAGGGGATCGCAACATCTGGATCTGTAAGCAGGAGCC 1113

QY 591 AAGTCCCGGGCGGAGACATAGTGTGCATGGACCGCTGTGGAGGAGATCTTGGAGCTGGCA 650

Db 1114 AAGTCCCGGAGGAGGATCATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGTG 1173
Qy 651 GCTGCAGACACCTCTTCCAGGGACAACAAGTGGTGGTCCAGAGTACATCCGAGAG 710
Db 1174 AACGGCAACCCGTGTGATGAAGAGCGCAAGTGGTGGTGCAGAAATATATTGAGCGG 1233
Qy 711 CCCTCTCATCTGTGACACCAAGTTGCGATCAGACAGTGGTTCCTCGTCACGGAAGTGG 770
Db 1234 CCCTCTCATCTTGGACCAAGTTGACCTCAGACAGTGGTTCCTCGTAACTGACTGG 1293
Qy 771 AACCCCTGACCATCTGGTTCTCAAGAGAGTTACTTGGGTTCTCAACTCAGCCCTTC 830
Db 1294 AACCCACTTACCGTGGTGTCTACCGCGACAGTATCCGCTTTTCCACGAGCCCTTC 1353
Qy 831 TCCTCGACAAGCTGCAGACGCGCATCCACCTGTGCAACAACGCGTCCAGAAGTACCTG 890
Db 1354 TCCTGAAGAACTGGACAACTCAGTGCACCTGTGCAACAATCCATCCAGAAGCACCTG 1413
Qy 891 AAGAATGATGTGGGCGCGAGCCCTGTGCTGCCGCAACAACAATGTGGACCAACAGG 950
Db 1414 GAGAACTCATGCGCATCCGCTGCTTCCGCCAGACAACATGTGGTGTAGCCAGAGG 1473
Qy 951 TTCAGGAGTACTGCAGCGCCAGCGCGTGGCGCGTGTGGGAGAGCTCATCTACCCG 1010
Db 1474 TTCAGGCGCCACCTGCAGGAGATGGGTGCCCAATGCTTGGTCCACCATCATCTGTCCT 1533
Qy 1011 TCATCAAGAAGGCCATCGCCAGCGCATGAGTGGCGCCAGGACCACTGGAGCTCGC 1070
Db 1534 GGCTGAAGATGCTGTGATTCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGG 1593
Qy 1071 AAGAACAGCTTTGAGCTACCGGGTGTGCTTGGTGGAGGAGTTCAGGCGCTGG 1130
Db 1594 AAGGCCAGCTTTGAGCTCATGGCGTGTGCTTGGTGGAGGAGTTCAGGCGCTGG 1653
Qy 1131 CTGATCGAGATCAATTCCAGCCCCACATGCGACCCCTGCAGCGCGGTTCAGGCGCAGCTG 1190
Db 1654 CTGATTGAGATCAACGCCAGCCCCAGGATGGACCTCCACAGCAGTCACTGCCCGCTC 1713
Qy 1191 TGTGCACAGTGCAGGAGACCACTCAAGTGGCC-----GTGGACGCGAGC 1238
Db 1714 TGTGCTGGCGTGCAGCTGACACCTTCGCGTGGTCAATTGACCGGATGTGGACCGCAAC 1773
Qy 1239 TGTGACATCGGCAACTTTCAGCTCTCTGTGGAGGAGCAGCCGT 1279
Db 1774 TGTGACACAGGACCTTTGAGCTCATCTATAGCAGCCCGT 1814

RESULT 12
ID ACN37881
XX ACN37881 standard; cDNA; 2553 BP.
XX ACN37881;
XX
XX
XX 18-NOV-2004 (first entry)
XX
XX Tumour-associated antigenic target (TAT) cDNA DNA324273, SEQ ID NO:1057.
XX
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.
XX Homo sapiens.
XX
XX WO2004030615-A2.
XX
XX 15-APR-2004.
XX
XX 29-SEP-2003; 2003WO-US028547.

XX 02-OCT-2002; 2002US-0414971P.
XX (GETH) GENENTECH INC.
XX Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
XX P-PSDB; ABM80420.
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX Claim 1; SEQ ID NO 1057; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT nucleic acid of the invention
XX
XX Sequence 2553 BP; 575 A; 716 C; 733 G; 529 T; 0 U; 0 Other;
XX
XX Query Match 22.9%; Score 421.4; DB 13; Length 2553;
XX Best Local Similarity 63.8%; Pred. No. 3e-80;
XX Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
XX
Qy 232 GGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA 291
Db 754 GGGATCGTGACAGGCCCTCCCTATGTGCAGAGAGACAGCAGCCCAAGAAACAGGAGAA 813
Qy 292 AGCTCAG-GGGCCTCCCGGGCAGCTTGTGGACATCGCGTGCAGAGTGTGCGAGGCTTAC 350
Db 814 AACCCAGTGTGGTGTCCCAGAGTGTGGATGAAGCTCTGTGCTGCGAGGAGTAC 873
Qy 351 CTGGGGCAGCTGGAGCATGAGACATCGACAGTGCAGCAGATGCCGTGGAGGACCTCCT 410
Db 874 CTTAGCAACTTGGGCCCAATGACATCGACAGGACCTGGAGGCCCGCTGTACTCTCAC 933
Qy 411 GAGGCGGAGTGGAGGACCTGACCAGCAGTACTCTCCCTCGTTCATGCCGATGCTTTC 470
Db 934 CCGAGGGCTGTCTCCCTCTTCTCCAGCGCTACTACCAAGTGTCCACGAAAGGGCAGAA 993
Qy 471 APTCCAAATTCAGAAATTAATCTTTTCGAGTCCCGAGGCTCTCTGTAATAGAATCAGCTT 530
Db 994 CTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTGCAGCAGCTGCAGGCC 1053
Qy 531 GTGAACCTTCAGACGACATTCACGGGCTCCGGAACATCTGGATATTAAGGCCCGGGCC 590
Db 1054 GTGGTACCCAGATAGACATGAAGGGGATCCCAACATCTGGATCGTGAAGCAGGAGCC 1113
Qy 591 AAGTCCCGGGCGGAGACATAGTGTGCATGGACCGCTGTGGAGGAGATCTCTGGAGCTGGCA 650
Db 1114 AAGTCCCGGGACGAGGACATCATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGTG 1173
Qy 651 GCTGCAGACCACTCTTTTCAGGAGCAACAAGTGGGTGTCAGAAATGATCGAGACG 710

Db 1174 AACGGCAACCCCGTGGTGTATGAAGGACGGCAGTGGGTGTCGAGAGTATATTGACGG 1233
Qy 711 CGCGTCTCATCTGTGTGACACAAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGG 770
Db 1234 CCCCTCTCATCTTTGGCACCAGTGTGACCTCAGACAGTGGTTCCTGGTAACTGACTGG 1293
Qy 771 AACCCCTGACCACTCTGTTCTACAGGAGAGTGTACTTGGGTTCTCAACTCAGCGCTTC 830
Db 1294 AACCCACTTACCCTGTGGTCTACCGGACAGCTATATCCGCTTTTCCACGACGCCCTTC 1353
Qy 831 TCCCTGACAAAGTGGACAGCGGCATCCACTGTGCAACAAACGCGTCCAGAGTACTCG 890
Db 1354 TCCCTGAAGAACTGGACAACTCAGTGCACTGTGCAACAACTCCATCCAGAGCACCTG 1413
Qy 891 AGAATGATGTGGCCCGACGCCCTGCTGCCCGCACACAACTGTGGACCAACCAAGG 950
Db 1414 GAGAACTCATGCGCATCGGCATCCACTGCTTCCGCCAGACAACTGTGCTTAGCCAGAGG 1473
Qy 951 TTCAGAGGTACCTGACAGCGCCAGGGCCGTGGCGCCGTGTGGGCGACGCTCATCTACCG 1010
Db 1474 TTCAGGCCACCTGACAGGAGATGGGTGGCCCCAAATGCTTGGTCCACATCATCGTGCCT 1533
Qy 1011 TCCATGAAGAGCCATCGCCCGCCATGAAGTGGCCCGCAGGACCAACGTCGGAGCCTCGC 1070
Db 1534 GGCATGAGGATGCTGTGATCCAGCACTTCAGACCTCCAGACACCGTGCAGTTCGG 1593
Qy 1071 AGAAGAGCTTTGAGCTCTACGGGGTGAATTCGTCTTGGGAGGAGTTCAGGCCCTGG 1130
Db 1594 AAGGCCAGCTTTGAGCTCTATGGCGTGAATTCGTGTTCGGGAGGACTTCCAGCCCTGG 1653
Qy 1131 CTGATCGAGTCAATTCAGCCCCACCACTGACCGCTCCAGCGCGTTCAGGCCCGAGCTG 1190
Db 1654 CTGATGAGATCAACGCCAGCCCCAGATGGCACCTCCACAGCAGTCACTGCGCCGCTC 1713
Qy 1191 TGTGCAGGTGACAGGAGACACATCAAGTGTGCGC-----GTGACCGCAGC 1238
Db 1714 TGTGCTGGGTGACGTGACCTGCGCGTGTGATGACCGGATGCTGGACCGGAAC 1773
Qy 1239 TGTGACATCGGCAACTTCGAGCTCTCTGTGGAGCAGCCGGT 1279
Db 1774 TGTGACACAGGACCTTTGAGCTCATCTATAGCAGGCCGT 1814

RESULT 13

AAC77214
ID AAC77214 standard; cDNA; 2881 BP.
XX
AC AAC77214;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2769 polynucleotide sequence SEQ ID NO:5537.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
XX
PN WO200058473-A2.
XX

PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimketa RA, Leach M;
XX
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB43005.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5; Page 4718-4720; 5507pp; English.
XX
CC AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease, to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 2881 BP; 640 A; 812 C; 840 G; 589 T; 0 U; 0 Other;
Query Match 22.9%; Score 421.4; DB 3; Length 2881;
Best Local Similarity 63.8%; Pred. No. 3.1e-80;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 232 GGGAGAGGCGGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA 291
DB 754 GGGATCGTGACAGGCCCTCCCTATGTGCAGGAGACAGCAGCCCAAGAAACAGGAGAAA 813
QY 292 AGCTCAG-GGGCTCCCGGGCAGCTTGTGGACATCGCTGCAAGGTGTGCCAGGCTCCTAC 350
DB 814 AACCCAGTGTGGTGTCCCGAGATTTGTGGATGAAGTCTGTGTGCGTGGAGGAGTAC 873
QY 351 CTGGGGCAGCTGAGCATGAGGACATCGACACATCGACAGATGCGCGTGGAGGACCTCACT 410
DB 874 CTTAGCACTTGGCCCATGACATCGACATCGACAGGACCTGGAGGCCCGCTGTACTCACC 933
QY 411 GAGGCCGAGTGGAGGAGCCTGACCCAGAGTACTACTCCCTCGTTTATGCGGATGCTTTC 470
DB 934 CCGAGGGGTGTGCTCTTCTCCAGCGCTACTCAAGTGTGTCCACGAAGGGGAGAA 993
QY 471 ATCTCCAATTCAGAAATTTACTTTTCGAGTCCAGGCTCTGCTGAATAGAACTGCTCT 530
DB 994 CTCAGGCACTCGACACTCAGGTCAGGTCAGGATCTGTGAGGACATCTCTGACGAGTGC 1053
QY 531 GTGAACCTCTCAGACGACATTCAGCGGCTCCGGAACATCTGGATTTATAAGCCCGGGCC 590
DB 1054 GTGGTACCCAGATAGACATGAGAGGGGATCGACATCTGGATCTGTGAGCCAGGAGCC 1113


```

QY      831  TCCTGGAAGGTGACAGCGCCATCCACCTGTGTGCAACAAAGCCCTCCAGAGTACCTG 890
Db      1823  TCCTGAAGAACCTGACAACTCAGTGCACCTGTGCAACAACTCCATCCAGAAGCACCTG 1882

QY      891  AAGAATGATGTGGCGCGCAGCCCTGCTGCGCGCACACAAATGTGGAGCCAGCACCCAGG 950
Db      1883  GAGAACTCATGCCATCGGCATCCACTGCTTCGCGCCAGAACAAATGTGGTCTAGCCAGAGG 1942

QY      951  TTCAGGAGTACCTGACAGCCCGAGCGCGTGTGGCGCGTGTGGGGCAGCGTCACTTACCCG 1010
Db      1943  TTCAGGCCCACTGACAGGATGGGTGCCCCAAATGCTTGGTCCACCATCATCGTGCCT 2002

QY      1011  TCATGAAGAAGGCCATCGCCACGCCATGAAGGTGGCCCGAGCAACGCTGGAGCCTGCG 1070
Db      2003  GGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGCG 2062

QY      1071  AAGAACAGCTTTGAGCTCTACGGGGTGAATTCGTCCTTTGGGAGGGAATTCAGGCCCTGG 1130
Db      2063  AAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCGTGTTTGGGGAGGACTTCCAGGCCCTGG 2122

QY      1131  CTGATCGAGATCAATTCAGCCCCACATGCAACCCGTCACCGCGGTCAAGGCCAGCTG 1190
Db      2123  CTGATTGAGATCAACGCCAGCCCCACATGGCAACCCCTCCACAGCACTCACTGCCCGGCTC 2182

QY      1191  TGTGCACAGGTGACAGGAGACACCAATCAAGGTGGCC-----GTGACCGCAGC 1238
Db      2183  TGTGCTGGCGTGCAGAGCTGACACCCCTGCGCGTGGTCAATTGACCGGATGCTGGACCGCAAC 2242

QY      1239  TGTGACATCGGCAACTTCGAGCTCCTGTGGAGGCAGCCGCT 1279
Db      2243  TGTGACACAGGAGCCCTTGAGCTCATCTATAGCAGCCCGT 2283

```

Search completed: September 24, 2005, 18:03:24
 Job time : 1125.21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:30:53 ; Search time 10402.7 Seconds
(without alignments)
10438.446 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgctgcagcagagccggc.....gcctccacaggggcccctccc 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	2241	100.0	101270	9 HS355C18	AL022327 Human DNA
2	136	6.1	374	6 CQ724916	CQ724916 Sequence
3	108.8	4.9	163783	10 AC119959	AC119959 Mus muscu
4	102.6	4.6	218249	2 AC097425	AC097425 Rattus no
5	102.2	4.6	125020	9 AF429315	AF429315 Homo sapi
6	101.6	4.5	773	11 BV054314	BV054314 S212P6049
C 7	79	3.5	125020	9 AF429315	AF429315 Homo sapi
8	76.6	3.4	197360	2 AC018829	AC018829 Homo sapi
9	76.4	3.4	145435	2 AC026685	AC026685 Homo sapi
C 10	76.4	3.4	185067	9 AC022382	AC022382 Homo sapi
C 11	76.4	3.4	189430	2 AC011610	AC011610 Homo sapi
12	76.4	3.4	191834	2 AC026196	AC026196 Homo sapi
C 13	76.4	3.4	216911	9 AC018809	AC018809 Homo sapi
14	76.4	3.4	222994	2 AC016947	AC016947 Homo sapi
15	73.2	3.3	155313	2 AC068315	AC068315 Homo sapi
16	72.4	3.2	1299	10 BC021404	BC021404 Mus muscu
17	71.6	3.2	5282	6 CQ842940	CQ842940 Sequence
18	71.6	3.2	5282	9 AK125875	AK125875 Homo sapi
C 19	71.6	3.2	224077	2 AC120010	AC120010 Mus muscu

C 20	71.6	3.2	266888	2 AC137881	AC137881 Mus muscu
21	71.4	3.2	252689	2 AC079433	AC079433 Mus muscu
C 22	69.8	3.1	7218	6 I66494	I66494 Sequence 14
23	68.6	3.1	209887	2 AC079424	AC079424 Mus muscu
C 24	68.2	3.0	1897	10 BC006830	BC006830 Mus muscu
C 25	67	3.0	64976	2 AC026256	AC026256 Homo sapi
C 26	67	3.0	267894	2 AC117950	AC117950 Rattus no
27	66.8	3.0	239130	2 AC079420	AC079420 Mus muscu
28	66.4	3.0	202083	2 AC023833	AC023833 Mus muscu
29	65.4	2.9	300695	2 AC079431	AC079431 Mus muscu
C 30	65	2.9	303091	2 AC084799	AC084799 Mus muscu
C 31	64	2.9	126543	2 AC143806	AC143806 Macaca mu
32	63.8	2.8	265537	2 AC087228	AC087228 Mus muscu
33	63.4	2.8	3281	2 AC143602	AC143602 Macaca mu
34	63	2.8	261449	2 AC095574	AC095574 Rattus no
35	62.6	2.8	63087	2 AC023446	AC023446 Homo sapi
36	62.4	2.8	736	11 BV106786	BV106786 PZA01640
C 37	62.4	2.8	840	8 CNS018NS	AL114464 Botrytis
C 38	62.4	2.8	134940	2 AC018939	AC018939 Homo sapi
C 39	62.2	2.8	141003	2 AC016463	AC016463 Homo sapi
40	62.2	2.8	161903	2 AC021996	AC021996 Homo sapi
C 41	62	2.8	250046	2 AC113592	AC113592 Mus muscu
42	62	2.8	298166	2 AC087563	AC087563 Homo sapi
43	61.8	2.8	132001	2 AC142984	AC142984 Macaca mu
44	61.6	2.7	128658	2 AC147014	AC147014 Medicago
45	61.6	2.7	207420	2 AC078884	AC078884 Mus muscu

ALIGNMENTS

RESULT 1
HS355C18/c

LOCUS

DEFINITION

HS355C18 101270 bp DNA linear PRI 05-JUN-2003
Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3
Contains the KIAA0027 gene, ESTs, STSs, GSSs and seven putative CpG islands, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL022327

AL022327.17 GI:5304851

HTG; CpG island; KIAA0027.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 101270)

Direct Submission

Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

On Jun 30, 1999 this sequence version replaced gi:5262834.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submissions corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep -----

Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

RP3-355C18 is from the library RPCI-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

This sequence is the entire insert of clone RP3-355C18. The true left end of clone RP5-898I4 is at 5390 in this sequence.

FEATURES

source	1. .101270 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="22" /map="q13.3" /clone="RP3-355C18" /clone_lib="RPC1-3"	repeat_region	8306. .8316 /notes="2.8 copies 4 mer GTGG 22% conserved"
		repeat_region	8399. .8414 /notes="3.2 copies 5 mer GAGGG 23% conserved"
		repeat_region	8840. .8851 /notes="2.4 copies 5 mer AGCTG 24% conserved"
		repeat_region	8939. .8958 /notes="2.2 copies 9 mer CCCCTCCCCA 40% conserved"
misc.feature	1. .1038 /note="match: STS: Em:AL022483" complement(1. .242) /note="AluJo repeat: matches 1. .237 of consensus" 358. .484 /note="L1M4 repeat: matches 829. .959 of consensus"	repeat_region	9125. .9135 /notes="2.2 copies 5 mer GTCCT 22% conserved"
		repeat_region	9246. .9256 /notes="3.7 copies 3 mer GGA 22% conserved"
		repeat_region	9263. .9274 /note="2.0 copies 6 mer CAGTGA 24% conserved"
		repeat_region	9373. .9383 /notes="2.2 copies 5 mer CTGA 22% conserved"
		repeat_region	9434. .9444 /notes="2.2 copies 5 mer CACGC 22% conserved"
		repeat_region	9616. .9631 /note="2.3 copies 7 mer GGGGATG 32% conserved"
		repeat_region	10067. .10076 /notes="2.5 copies 4 mer GGCA 20% conserved"
		repeat_region	10245. .10254 /notes="2.5 copies 4 mer CCTT 20% conserved"
		repeat_region	10496. .10531 /note="2.0 copies 18 mer GCGAGGACAGGACAGG 63% conserved"
		repeat_region	10498. .10531 /note="2.8 copies 12 mer CAGGGGACGGGA 59% conserved"
		repeat_region	10745. .10756 /notes="2.0 copies 6 mer GCTGGG 24% conserved"
		repeat_region	10889. .10903 /notes="2.1 copies 7 mer CCGGCC 30% conserved"
		repeat_region	11166. .11193 /note="3.5 copies 8 mer TGTGTGCA 33% conserved"
		repeat_region	11177. .11211 /notes="5.8 copies 6 mer GTGCAT 38% conserved"
		repeat_region	11250. .11261 /notes="2.0 copies 6 mer ATGGGT 24% conserved"
		repeat_region	11273. .11282 /notes="2.5 copies 4 mer TGCC 20% conserved"
		repeat_region	11293. .11309 /notes="2.8 copies 6 mer TGTGTA 25% conserved"
		repeat_region	11311. .11411 /note="2.5 copies 40 mer TGATGGGTGTCATGGGTATGGGTGCGTGTGCTGTGCCT 202% conserved"
		repeat_region	11323. .11334 /note="2.0 copies 6 mer ATGGGT 24% conserved"
		repeat_region	11346. .11355 /notes="2.5 copies 4 mer TGCC 20% conserved"
		repeat_region	11368. .11379 /notes="2.0 copies 6 mer ATGGGT 24% conserved"
		repeat_region	11391. .11400 /notes="2.5 copies 4 mer TGCC 20% conserved"
		repeat_region	11431. .11683 /note="7.0 copies 36 mer TGCATGTGTATGGGTGTGTATGTCATGGGTGCGTG 108% conserved"
		repeat_region	11451. .11490 /notes="2.9 copies 14 mer TGTGCATGGGTGCG 71% conserved"
		repeat_region	11452. .11477

		/note="3.2 copies 8 mer GTGCATGG 29% conserved" 11485. .11498 /note="2.3 copies 6 mer TGGGTG 28% conserved" 11513. .11523 /note="2.8 copies 4 mer TGCC 22% conserved" 11521 /note="Tandem repeat. Assembly not confirmed by digest" 11531. .11562 /note="5.3 copies 6 mer TGTGCA 32% conserved" 11536. .11635 /note="6.2 copies 16 mer GTGTGCATGTGTATGG 82% conserved" 11595. .11604 /note="5.0 copies 2 mer TG 20% conserved" 11679. .11690				
	Query Match	100.0%; Score 2241; DB 9; Length 101270;				
	Best Local Similarity	100.0%; Pred. No. 0;				
	Matches 2241; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	TGCTGTCAGCAGAGCCGGCTGGGATCTCTACGCTGCACGCCATTTGTCTCTGTATGGTT	60			
Db	60138	TGCTGTCAGCAGAGCCGGCTGGGATCTCTACGCTGCACGCCATTTGTCTCTGTATGGTT	60079			
Qy	61	GCCTCACTCATCCCATCTCTGGCTGGTGTGAACCTTGAATGCTGGGCATCAATAAAGACT	120			
Db	60078	GCCTCACTCATCCCATCTCTGGCTGGTGTGAACCTTGAATGCTGGGCATCAATAAAGACT	60019			
Qy	121	TTTTTCTTCGACGTAGGTTGGTAAATAATGAGATGCCGTACCTCCTCTGCAGCATCAA	180			
Db	60018	TTTTTCTTCGACGTAGGTTGGTAAATAATGAGATGCCGTACCTCCTCTGCAGCATCAA	59959			
Qy	181	GAGGGACATCATTTGACTATCACAGCTGCACCTACGACACAGATGCTGAACCATACGACAAA	240			
Db	59958	GAGGGACATCATTTGACTATCACAGCTGCACCTACGACACAGATGCTGAACCATACGACAAA	59899			
Qy	241	GACAGCTCTCTTCAACCAACCAAGTGAACCGCGCGCTTTGACGCTGGGGAGCCGGCAA	300			
Db	59898	GACAGCTCTCTTCAACCAACCAAGTGAACCGCGCGCGCTTTGACGCTGGGGAGCCGGCAA	59839			
Qy	301	GGGTGGTGGGGAGGAGCTTCTGCAGCCATATAGGACCTTCGGTGGCTGGTCAGTGGCCAC	360			
Db	59838	GGGTGGTGGGGAGGAGCTTCTGCAGCCATATAGGACCTTCGGTGGCTGGTCAGTGGCCAC	59779			
Qy	361	CAGTCACCTCTCTGGGTGCATCCACGAGCTCTCTGGTCTAAGGCCCTGGCCAGAAATCACTCG	420			
Db	59778	CAGTCACCTCTCTGGGTGCATCCACGAGCTCTCTGGTCTAAGGCCCTGGCCAGAAATCACTCG	59719			
Qy	421	GTGCCACCCCAACCCACAGACCCCTGTGCGCTTTGCTCTGTGTCTCTGGGTGAATCCGG	480			
Db	59718	GTGCCACCCCAACCCACAGACCCCTGTGCGCTTTGCTCTGTGTCTCTGGGTGAATCCGG	59659			
Qy	481	GGCCCCAGAGCTCCCTCTCAGGGCACAAGGCCAAGATGGGCTGATCTGGGGGCTGC	540			
Db	59658	GGCCCCAGAGCTCCCTCTCAGGGCACAAGGCCAAGATGGGCTGATCTGGGGGCTGC	59599			
Qy	541	CACCGGGCTTTGGGTGTCTAGAGGGGCTGTGGGACCCACAGGGGAAGAGTGCCTGCCCC	600			
Db	59598	CACCGGGCTTTGGGTGTCTAGAGGGGCTGTGGGACCCACAGGGGAAGAGTGCCTGCCCC	59539			
Qy	601	CAGCACCGGGCACTCCCGCGCTCTCCCGCGCTCTCAGCCCTGCACAAAGCAAGCTTGGCCAC	660			
Db	59538	CAGCACCGGGCACTCCCGCGCTCTCCCGCGCTCTCAGCCCTGCACAAAGCAAGCTTGGCCAC	59479			
Qy	661	ACCTTTCCACCAAGGCCCAAGGCTAGAGGCTGCCACAGGACCCAGGAGTGGGGACCTGTG	720			
Db	59478	ACCTTTCCACCAAGGCCCAAGGCTAGAGGCTGCCACAGGACCCAGGAGTGGGGACCTGTG	59419			
Qy	721	CTGAGGAGGGGTCCCGGAAGAGGGGCTTCCTCTGGGCACAGAGGTCCCTCTCAGCAGGCGCA	780			
Db	59418	CTGAGGAGGGGTCCCGGAAGAGGGGCTTCCTCTGGGCACAGAGGTCCCTCTCAGCAGGCGCA	59359			
Qy	781	GGTCCGGCTGCCTCAGCACAGTGTGGGGCGGAGGTGCAGGACAAAGGTTCCCTTCCGCAACC	840			

Qy	1921	GC	CCCATCGGCTATGAGGGCGGAAGTGGGGCGGGTTTGGGGAGCCTCGGTGGCCCTGGCC	1980
Db	58218	GC	CCCATCGGCTATGAGGGCGGAAGTGGGGCGGGTTTGGGGAGCCTCGGTGGCCCTGGCC	58159
Qy	1981	CCAT	CTCCAGTCCCGAGTCTCGGCTCGGACAGATAGGGCGAGGCTGTGCTGTCTTTCAG	2040
Db	58158	CCAT	CTCCAGTCCCGAGTCTCGGCTCGGACAGATAGGGCGAGGCTGTGCTGTCTTTCAG	58099
Qy	2041	AAGACT	TCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACGAGCT	2100
Db	58098	AAGACT	TCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACGAGCT	58039
Qy	2101	GC	AGCAGGAGCAGCAGAAAGCCAGGCCAGGAGGAGGAGGCGGGAGCAGCGACC	2160
Db	58038	GC	AGCAGGAGCAGCAGAAAGCCAGGCCAGGAGGAGGAGGCGGGAGCAGCGACC	57979
Qy	2161	TG	AGCAGCAGGCAAGGTGCGCTGGCGGGCGGCGAGGCGAGTGCACGAGCCACGAGCTCGG	2220
Db	57978	TG	AGCAGCAGGCAAGGTGCGCTGGCGGGCGGCGAGGCGAGTGCACGAGCCACGAGCTCGG	57919
Qy	2221	GGCCT	CCACAGGGGGCCTCCC	2241
Db	57918	GGCCT	CCACAGGGGGCCTCCC	57898
RESULT 2				
LOCUS	CQ724916	374 bp	DNA	linear
DEFINITION	Sequence 10850 from Patent WO02068579.			PAT 03-FEB-2004
ACCESSION	CQ724916			
VERSION	CQ724916.1	GI:42285773		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1			
TITLE	Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W. Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof			
JOURNAL	Patent: WO 02068579-A 10850 06-SEP-2002; PE Corporation (NY) (US)			
FEATURES	Location/Qualifiers			
source	1..374 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"			
ORIGIN				
Query Match	6.1%	Score 136;	DB 6;	Length 374;
Best Local Similarity	100.0%;	Pred. No. 2.8e-16;		
Matches 136;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	2040	GA	AGACTTCGCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACGAGC	2099
Db	239	GA	AGACTTCGCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACGAGC	298
Qy	2100	TG	CAGCAGGAGCAGCAGAAAGCCAGGCCAGGAGGAGGAGGCGGGAGCAGCGAC	2159
Db	299	TG	CAGCAGGAGCAGCAGAAAGCCAGGCCAGGAGGAGGAGGCGGGAGCAGCGAC	358
Qy	2160	CT	CAGCAGCAGGGCAAG	2175
Db	359	CT	CAGCAGCAGGGCAAG	374
RESULT 3				
LOCUS	AC119959	163783 bp	DNA	linear
DEFINITION	Mus musculus chromosome 15, clone RP24-467H19, complete sequence.			AC119959
ACCESSION	AC119959			

AC119959.8	GI:50540805
HTG.	
Mus musculus (house mouse)	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
1 (bases 1 to 163783)	
Birren, B., Nusbaum, C. and Lander, E.	
Mus musculus chromosome 15, clone RP24-467H19	
Unpublished	
2 (bases 1 to 163783)	
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazarro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.	
Direct Submission	
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
3 (bases 1 to 163783)	
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Mlenga, V., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.	
Direct Submission	
Submitted (26-MAY-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
4 (bases 1 to 163783)	
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,	

McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
 O'Connor, P., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
 Totham, K., Travers, M., Vasiliev, H., Venkataranan, V. S., Viel, R.,
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
 Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-JUL-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 23, 2004 this sequence version replaced gi:47679219.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L25717

Center clone name: 467_H_19

FEATURES

source

Location/Qualifiers

1..163783
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="15"
 /map="15"
 /clone="RP24-467H19"
 /clone_lib="RPC1-24 Male Mouse BAC"

misc_feature

1..4
 /note="clone boundary
 clone end:SP6
 site:MboI"

repeat_region

complement(169..204)
 /rpt_family="ORR1B1"
 818..994
 /rpt_family="ORR1D"
 1312..1527
 /rpt_family="B3"
 1573..1645
 /rpt_family="ID3"
 1665..1793
 /rpt_family="B1F"
 1794..1959
 /rpt_family="B3"
 complement(2256..2658)
 /rpt_family="L1"
 complement(2659..2833)
 /rpt_family="B2_Mm2"
 complement(2834..3302)
 /rpt_family="L1"
 complement(3311..3448)
 /rpt_family="B1_MM"
 complement(3449..4218)
 /rpt_family="L1"
 4214..4557
 /rpt_family="Lx6"
 5120..5279
 /rpt_family="CT-rich"
 complement(5297..5423)
 /rpt_family="B1F"
 complement(5451..5524)
 /rpt_family="PB1D9"
 6016..6494
 /rpt_family="BLTR11A"
 complement(6678..6797)
 /rpt_family="B1_MM"

repeat_region complement(6810..6945)
 /rpt_family="B1_MM"
 7008..7148
 /rpt_family="B1_MM"
 repeat_region complement(7763..7941)
 /rpt_family="B3"
 repeat_region complement(7960..8105)
 /rpt_family="B1_MM"
 repeat_region complement(8121..8300)
 /rpt_family="B3A"
 8486..8625
 /rpt_family="B4A"
 8618..8788
 /rpt_family="B1_MM"
 repeat_region 8804..9102
 /rpt_family="L1MB2"
 9162..9217
 /rpt_family="PB1D7"
 repeat_region complement(9339..9460)
 /rpt_family="B3"
 repeat_region complement(9598..9748)
 /rpt_family="B3"
 repeat_region complement(9755..12502)
 /rpt_family="L1_MM"
 12503..12992
 /rpt_family="IAPLTR2_MM"
 12994..13717
 /rpt_family="IAPLTR2_MM-int"
 13718..13836
 /rpt_family="IAPLTR2_MM-int"
 13832..14083
 /rpt_family="IAPLTR2_MM-int"
 14083..15904
 /rpt_family="IAPLTR2_MM-int"
 15905..16394
 /rpt_family="IAPLTR2_MM"
 complement(16395..20012)
 /rpt_family="L1"
 repeat_region complement(20013..20157)
 /rpt_family="B1_MM"
 20192..20290
 /rpt_family="B2_Mm2"
 20410..20553
 /rpt_family="B1_MM"
 20586..20777
 /rpt_family="B2_Mm2"
 21012..21082
 /rpt_family="(CCAA)n"
 21192..21294
 /rpt_family="PB1D10"
 21301..21342

Query Match 4.9%; Score 108.8; DB 10; Length 163783;

Best Local Similarity 84.7%; Pred. No. 1.9e-11;

Matches 122; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 122 TTTTCTTCCAGTCTAGGTTGGTAAAAAATGAGATGCCGTACCTCTCTTGGACCATCAAG 181
 Db 118788 TCTTCTTCCAGTCCAGGTTGGTAAAGACGAGATCCATACCTCTCTGACTATCAAG 118847
 Qy 182 AGGAGATCATTTGACTATCAGCCTGACCTACGACCATGCTGACCATCAAG 241
 Db 118848 AGGGATGTTGGGACTACCAAGCTGACCTGACCTGACCATGCTGACCATCAAG 118907
 Qy 242 ACAGCCTCTTCCACCAAGGTTG 265
 Db 118908 ACAGCATTCATCAACCAAGGTTG 118931

RESULT 4

AC097425

LOCUS

DEFINITION

AC097425 218249 bp DNA linear HTG 10-MAY-2003
 Rattus norvegicus clone CH230-137J2, *** SEQUENCING IN PROGRESS

***, 2 unordered pieces.
AC097425 GI:30521318
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 218249)
Muzny, D., Maric, M., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschrocks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flegg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Loulseg, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasaana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 218249)
Worley, K.C.
Direct Submission
Submitted (18-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 218249)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23268852. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GIGR
Center clone name: CH230-137J2
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 197170 bases at least Q40
Consensus quality: 202160 bases at least Q30
Consensus quality: 205125 bases at least Q20
Estimated insert size: 209885; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 216852: contig of 216852 bp in length
* 216853 216952: gap of unknown length
* 216953 218249: contig of 1297 bp in length.

FEATURES
source
1. 218249
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-137J2"
/notes="wgs_contig"
misc_feature
111665..1113309
/notes="wgs_contig"
misc_feature
143616..145372
/notes="wgs_contig"

ORIGIN
Query Match 4.6%; Score 102.6; DB 2; Length 218249;
Best Local Similarity 80.5%; Pred. No. 3e-10;
Matches 120; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
YQ 1434 TGCTTAACAGCCACCTGCTCTCGAGATCGGGCTGCGTGAACATGCGGAGCTGCCT 1493
Db 11483 TGCCTGATGCATGCTCTCTCGAGATGGAATCTCGCTGAACATGAGAGTCTGCCT 11542
YQ 1494 GGTACGTCCCGCCAAACCCGACTCTCTTCCACGCTCTACAGCTCTGCACCGAGA 1553
Db 11543 GGTACGTCCAGCCACCCGAAACCTCTTCCACGCTCTATGCGCTTGCACTGAGA 11602
YQ 1554 GTGACGACGAGGAGTCTCTGGGTAAGTGA 1582
Db 11603 GCGAGAAGCAGGAATCTTGGGTGAGTGA 11631

RESULT 5
AF429315

LOCUS AF429315 125020 bp DNA linear PRI 18-JAN-2002
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S.,
Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A.,
Potter, N.T., Ross, C.A. and Margolis, R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE
2 (bases 1 to 125020)
Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.
Direct Submission
AUTHORS
TITLE Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
JOURNAL
FEATURES
source
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(<36507..>36887)
/gene="JPH3"
/note="synonym: JP3"
complement(<36507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSGGRFNFDGSGYCGWEDKAHGHVCTGPKGQGYTGWS
HGFEVLGVYTPSGNTYGGTWAQGRHIGLESKRWYKGEWTHGPKRGYVRECA
NGAYKEGTWSNGLQDGYGTETSDG"
ORIGIN
Query Match 4.6%; Score 102.2; DB 9; Length 125020;
Best Local Similarity 11.7%; Pred. No. 4e-10;
Matches 188; Conservative 711; Mismatches 689; Indels 16; Gaps 6;
Qy 388 CTCCTGCTTAAGCCGTGGCCAGAACTCACTCGGTGCCCCACCCAGCCAGACCCCTG 447
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50557 CTGCTGCTGGGGCCACACCTCTGRTGKBCNSKSHSSCHRRPTWTDYRKYTYCCOYT 50616
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 448 TGCCCTTTGCTCTGCTCTGCGTGAATCCGGGGCCCCAGAGCTCCCTCTCAGGGCA 507
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50617 AYMNBVHYCYDBCWSGVWSYGGWRGMKBWBMBBSMMGKMDGTDKDYVCVYSM 50676
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 508 CAGAGGCCAAGATGGGGCTGACTGGGGCTGCCACCGGGCTTGGGCTTGAGGGGCT 567
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50677 SVDVVRWRBDSWGSBSHYMRGVTGVMRVWVGBCDTHVTYDGHGHSNMGKSCYS 50736
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 568 GTGGACCCAGGGGAGAGGTGCGCTCCCCCAGCACCGGGCACTCCCGCGCTCCC 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50737 DBGKCBGSKMCTGSKTKYCYGYKYSKSDBSYYSBSHYBWSGMYMCCRKLSYAS 50796

Qy 628 CGGCTCCAGCCCTGCACAAAGCAGCTTGCCACACCCCTTCCCAACCAAGCCCGAGGGCTGA 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 50797 SYDVRYYKBTGMRYRVRSCTSYYYKYBSMSVARSKSKGHKVRBBSMKSRKDMSTSSMW 50856
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 688 GGCCTGCCAGCAGCAGGGGTGGGGACCTGCTGAGGAGGGGTCCCGAAGAGGGGC 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 50857 RSMKDRKCSRSHSHYSKSMGKWTSDVRCVACSBMYKVMHRKHVKKDHSBWSGKSHM 50916
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 748 TTCCTTGCCACAGAGG---TCCCTCTCAGCAGGCCAGGTGCGGCTCGCTCAGCACAGTGT 804
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 50917 KKGAWRVRSMYRSMKCSVHDSSMVDKSRRRMKSGAMRMKCYCTSSMRSTKSRSMKSSW 50976
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 805 GGGG---CGAGGTGAGGACAAGTTCCTTCGCGACCTTAATACCCAGGTCAGGCC 860
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 50977 RWGSKCYCYGYWCTKMRSMGCSYSTGSSMKCMGKYCAGGRSSRYCKSGSRMSMAG 51036
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 861 AGCCGACGCGCTGCTAGTGCCAACTAGCCCTCTTCAAAGACCCCGTGTGCAGAGCCAGTC 920
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51037 GSMBSMSMGKMTSGMSRCARSRMSAYSCSKCKSKRCSGCMRCSWYCWYKMRVGS 51096
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 921 AGCGTCGCTGTGTTTCTGGGGCCAGCCACTTTTTCCTTAACAGGGTGACAAACAG 980
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51097 WRMYKCMSTYKTKYKGSATKSMAMVYSCMRMRKGTKSSAWTSMCCSKCARAASGWYK 51156
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 981 ACATCGGGGTGCGGGAGCTTCAAGATATGCTCTCTAATGCGCGGTGCTACTCGCGTGT 1040
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51157 CMMKGGCTGGGKGSASTWYMMCCSRAC--KYTTTRKSSRGRWGKATRTTRTSSSRM 51214
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1041 GGTGGCAGTGTGCGGATGCTGCGTTCGGGCGCTGCGAGCGCTCCCTCTCTCTCT 1100
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51215 SWYTYAMMRRTWTAGKKWTKTTWTKKSKMRMTSRWNGMTTWTYTKTTTDDV 51274
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1101 GGGTCTGAGGTGTGGGACACCTGTGCGCCGCTTGGGGGCTCGAGGAGCCCGCAGTCG 1160
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51275 YKKKSKMKSKWYTTCTKYTYKTYTYKTYKTYKTYKTYKTYKTYKTYKTYKTYKTYK 51334
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1161 CCACCGCCCTCGGTGCGCTGAGCTGAGCTCCCATTTCCCTCGGGGCTGCGACGCCCTC 1220
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51335 CCMGAGGSGWRGAKYSMMSTSKYGRWTQWRSCTWYCTSCMTRSTCTCMGCTCTCTSMS 51394
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1221 CACTGCTCAAGGATGCTGAGCTGCGCTGCGCTGCACATCCATGCTCCACCGGCTGGAG 1280
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51395 KRGTYSRRCVAYWSKCSYSYSCMRCCWYSYKSRSTRRKWTYSWGRYKTYKRWMS 51454
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1281 GGGTGTTCGGGGCCCTGGGGGACCGTATATCTGGGAGGGGAGGATCCGCCCTAGGG 1340
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51455 MSAGMYKGSKWYTYWYWTATKKYA-----GKMRGACGSKTITTCATRTTYM 51509
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1341 GTGCTGTGTGTCCTCCCATAGCCAGGCAAGTCTGAGGTGAGCGCCGCTCTCTCGCTTGT 1400
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51510 STSMMSKMWGTSWYSYCAKCTCGMCTCCCMRKRKYKSGMTKMCWCMGYSTYSM 51569
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1401 GAGCGAGCAGCTGAGCTGAGCTGCCCTCTCTGCTTAACACGCCACTGTCTCTGGAGA 1460
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51570 RMRSTKSGMYTRSMRSMSTKWSYATGMKYTYRRMMWMMWMMWMMWMMWMMWMMWMMW 51629
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1461 TCGGGCTGTGCGTGAACATCGGAGCCTGCGCTGCTACGCTCCCGGCAACCCCGACTCCT 1520
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51630 WYSNWSNRWATRRATWKGKAGYKRWGGTGTWTTMYAICWYCMYMWANGSSCMACCCYYC 51689
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1521 TCTTCCACGCTGCTACAGC-CTCTGACCCGAGAGTGACGACGAGGAGTTCCTCGGTAA 1579
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51690 ASMTGGYMSRAKSKWMMCCYCCCMSSMMCYYYCMGYSYKTYTYSYACTMSCKSAY 51749
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1580 TGAGGAGCGCAGAGGCCCCAGTCTGCTCAGCAGAGAGGCTTCTAGAAAGATCCCCCT 1639
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51750 SCMYMSKACKSKYCAKSGCYKKYWGSYYSYSGSRM-WYTSYYSYCMYMSMMWMSYMS 51808
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1640 GGTGCTCAGACACACTGATGGGCGAGGCTCTGAGGATAGAGGACCGGGAGAGGCTCCC 1699
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 51809 YSKSCYYSWRRKMCWCSWYRSARCTWRSWTSCYWSRRSTGWRRSKGWGCWYRRS 51868
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
QY 1700 ATGTCATGTCATGGCAGTACAGAGCCAGGCGGCCCGGAGGAGGGCGGGCAGTCAA 1759
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51869 CCM5WTKYCKGKKSWSMKCATTTGKMSCKARKSMSCYCGKMCRRASCWGCKMSMTKCM 51928
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1760 GGAGTGTGTTGCTAGGCTTGTGACAGCGCGGCGGAGCAGTGGGGCAAGCCG 1819
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51929 RSYTSTCTCYGCMYKWKYKRRSSYKCYKKSMTKYSMRGMSCYRGRSYMYSYR 51988
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1820 GCTCCGTGGTGGAGGCGGAGGGGTGCAGCTGGACGCGCCGCGAGTCACAGAGACACTCG 1879
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51989 WRKMCWMTWKSKGKSYNSNSWMMWKGTSYWCCLASSYAMSKRMATSWGCCYMS 52048
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1880 GGAGAGCGGAGGAGCAGACTGGGGGTGTGGAGCAGCGCTGGGCCCCATCGCTATGAGGG 1939
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52049 YRSASWCGSTGSMRSYKSTWRSWMSMRGSMYSYMRGKGCASWSYRKKSKRGRS 52108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1940 CGGGAAGTGGGGGGTGGGGAGCTCCGTGGCCCTGGGCCCA 1983
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52109 MGRGSMTKSKGTGKGMWRGSGWGSYKWKVRCAMKCYCYW 52152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
BV054314/c
LOCUS BV054314 773 bp DNA linear STS 31-MAY-2003
DEFINITION S212P60493PES.TO CZECHII/Ei Mus musculus STS genomic, sequence
tagged site.
ACCESSION BV054314
VERSION BV054314.1 GI:31170109
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 773)
AUTHORS Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
TITLE The mosaic structure of variation in the laboratory mouse genome
JOURNAL Nature 420 (6915), 574-578 (2002)
MEDLINE 22354684
PUBMED 12466852
COMMENT Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 773
Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129SI/SvimJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the MGSV3 C57BL/6J assembly and SNP
detection was carried out by SSAHA-SNP. 225,000 reads were
annotated
as SNPs and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.
FEATURES
source
1. .773
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/Ei"
/db_xref="taxon:10090"
/map="4 15 34-720 89771213-89771899"
/clone_lib="CZECHII/Ei"
<1. .>773

STS
ORIGIN
Query Match 4.5%; Score 101.6; DB 11; Length 773;
Best Local Similarity 80.4%; Pred. No. 1.4e-09;
```

```
Matches 119; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1435 GCCTAACAGCCCACTGTCTCTGAGATCGGGCTGTGCGTGAACATCGGAGCCCTGCCCTG 1494
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 727 GCCTGATACACTGCTCTCTCTGGAGATTGGACTCTGCCTGAACATGAGAAGTCTGCCCTG 668
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1495 GTACGTCGCGGCCAACCCCGACTCTCTTCCACGCTGTACAGCTCTGCACCGAGAG 1554
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 667 GTACGTCGAAGCCCAACCCGAACACCTCTTCCACGCTGTATGGCCTTGCACGTAGAG 608
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 1555 TGAGCAGCAGGAGTCTCTGGTAAGTGA 1582
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 CGAAGAAGCAGGAATCTTGGGTGAGTGA 580
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
AP429315/c
LOCUS AP429315 125020 bp DNA linear PRI 18-JAN-2002
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AP429315
VERSION AP429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.B.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
source
1. .125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/notes="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(<36507..>36887)
/genes="JPH3"
/notes="synonym: JP3"
complement(<36507..>36887)
/genes="JPH3"
/product="junctophilin 3"
complement(<36507..36887)
/genes="JPH3"
/notes="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSGRFRNPDGGSYCGWEDKKAHGHVCTGPKGGEYTGWS
HGFEVLGVTPSGNTYQGTWQAQKRGHIGLESKGVVYKGEWTHGFKRGYGVRECA
NGAKYEGTWSNGLQDGYGTETVSDG"

ORIGIN
Query Match 3.5%; Score 79; DB 9; Length 125020;
Best Local Similarity 11.8%; Pred. No. 1.4e-05;
```

Matches	116;	Conservative	445;	Mismatches	410;	Indels	11;	Gaps	5;
Qy	443	CCCTGTGCCCTTGTCTCTGTGTCTCTGGGTGAATCCGGGGCCCCAGAAAGCTCTCCCTCTCA	502						
Db	17864	CMKAGCWMYTGKSKKKKKYASKRGRKGTYYWKRKMSAWSRKWRWYASMMRCAGMM	17805						
Qy	503	GGGCACAGAGCCCAAGATGGGCTGACTGGGGCTGCCACCGGCTTTGGTGTCTGAG	562						
Db	17804	GAACGASRSMKCCWGRYSAGSKSRGGTGYMKKGGKGSMSKKWGSSTSRRRGS	17745						
Qy	563	GGGCTGTGGACCCCAAGGAAGAGTGCCTCCCCCCCAGACACGGGCACATCCCCCGCC	622						
Db	17744	SAKSCSYMMGSMCMSCMSMAKSYMMCYYMYRMSYMSYYKYKYSCHMGSTSYSYSCC	17685						
Qy	623	TCCCCGGCTTCAGCCCTGCACAAAGCAGCT-----TGCCACACCTTCCCAACAAGCC	678						
Db	17684	WKMSWGSYKCMKYISGMSYSYSGSYSTSKKYKCSWSMYKCKTSKYRKRYSY	17625						
Qy	679	CAGGGCTGAGGCTGCCCAGGACGCGAGGTGTGGGACCTCTCAGGAGGGGTCCGG	738						
Db	17624	YWGGRKAKKKYACGRRRMSYWKCAKWMYSYCCWYTCMTYYTKYSSTCYKRG	17565						
Qy	739	AAGAGGGCTTCCTTGCCA--CAGAGTCCCTTCAGCAGGCAGGTGGCGTGCCTCAG	796						
Db	17564	GYWGSKTCYSAGKSRSMYCYMNRSSKSSSSMSMARSSWCMWGAGYRRSRKRSAGWA	17505						
Qy	797	CACAGTGTGGGCGAGGTGCAGGACAAGTTCCTTCGCACTAATATCCCAAGGTCA	856						
Db	17504	GWRSSGKRSTGMRACSKTSGYSTGRSMKKKGYSKYSRGMKGKKTCTYCMWKKYK	17445						
Qy	857	GGCAGCGCCAGCGCTGTAGTGGCAACATGGCCCCCTTCAAAGACCCCGTGTGCAGAGCC	916						
Db	17444	RKTSMCWYYMKSNGYK--RYRCCMKKKGTGYRGHSSKKSGYKSRGMRGSGSYTSC	17388						
Qy	917	AGTCAGCGTGCCTGTGTTTTCTGGGGCCAGCCACTTTTTCTTAAACAGGTTGACAAA	976						
Db	17387	WKS CWGYSMWKMYMSYKKRRSRMBSMSKSGWRGAYGRGCYSSSMWSRKRRSKCYSY	17328						
Qy	977	CCAGACATCGGGTGGCGGACCTTCACGATATGCCTCTCTAATGGCGGTGTACTCGC	1036						
Db	17327	KSYKGRGKMKMGWMMKRSKYWSSMKKMSWSKCYSTKSYKSGRRKSKGWRSTKSK	17268						
Qy	1037	GTGTGTGGGCAGTGTGGCGCATGTGCCTCGCTCCGGGCCCTGCAGCGCTCTCCCTC	1096						
Db	17267	AKSSMRMAGSKCTYGSYYNSNNRNRNMKTCGNMYRRS--RAMNGNAAAGCTTCCCC	17209						
Qy	1097	TCCTGGGCTGTAGTGTGGGACACCTGTGGCCCGTTTGGGGGCTCGAGAGGAGCCCCA	1156						
Db	17208	ANTNGGGGGAAGAGCGCGSASRASCYKGRMSKSCYRSGTTRRRCMKSKRRGRSGKSMG	17149						
Qy	1157	GTCCCCACGCCCTCGTGCCTTCAGCTCAGCCTCCCATTTCTCTGGGGCTGCACGCGC	1216						
Db	17148	MTGRSGGKTSYSAKSGRGCTYCWGGRKGRKMSSRKRMKYKSYRRKRWMTCKMCMY	17089						
Qy	1217	CCTCCACTGCCTCAAGATGCGTGTGAGCTCGCCTGCACATCCATGCTCCACCGGCTG	1276						
Db	17088	SMYMYCRSMCMCMCKSCCGCYSMGMSYSYSGKYSWGMSKYWRSYYSKRSTSKAWR	17029						
Qy	1277	GAGGGGTGTTCTGGGGCCCTGGGGACCGTATATCTGGGGAGGGGAGGAGATCGGCCT	1336						
Db	17028	SSKRGMTGGRYKGGGRSYKGGGYSYKGGGSGWGG--GGKSTRGSGAGKKSYSKXMSCCAR	16970						
Qy	1337	AGGGGTGCTGTGTGTGCCCATAGCAGGCAAGTCTGAGGTGGACGCCCGTCTCTCGCC	1396						
Db	16969	YKSWSYKCYSRWMCWSSYYYCWGGKCMSSCWSSMRKSKGGSWAKGWVWDKGSVTDK	16910						
Qy	1397	TTGTGAGCAGGCGAGCTGAG	1418						
Db	16909	SDKMBRSBSKVKSXMSMRG	16888						

RESULT 8
AC018829

LOCUS	AC018829	197360 bp	DNA	linear	HTG 03-FEB-2000			
DEFINITION	Homo sapiens chromosome 3p clone RP11-402P11, WORKING DRAFT SEQUENCE, 33 unordered pieces.							
ACCESSION	AC018829							
VERSION	AC018829.3	GI:6862654						
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.							
SOURCE	Homo sapiens							
ORGANISM	Homo sapiens (human)							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
AUTHORS	Li, L., Hu, S., Dong, W., Zhang, X., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Wang, X., Feng, X., Yu, J., and Yang, H.							
TITLE	Chromosome 3p genomic sequence							
JOURNAL	Unpublished							
REFERENCE	2 (bases 1 to 197360)							
AUTHORS	Li, L., Hu, S., Dong, W., Zhang, X., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Wang, X., Feng, X., Yu, J., and Yang, H.							
TITLE	Direct Submission							
JOURNAL	Submitted (21-DEC-1999) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China							
COMMENT	On Feb 3, 2000 this sequence version replaced gi:6684183. * NOTE: This is a 'working draft' sequence. It currently * consists of 33 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.							
	1	2243:	contig of 2243 bp in length					
			gap of unknown length					
	2244	4381:	contig of 2138 bp in length					
			gap of unknown length					
	4382	6628:	contig of 2247 bp in length					
			gap of unknown length					
	6629	8467:	contig of 1839 bp in length					
			gap of unknown length					
	8468	10758:	contig of 2291 bp in length					
			gap of unknown length					
	10759	13438:	contig of 2680 bp in length					
			gap of unknown length					
	13439	16301:	contig of 2863 bp in length					
			gap of unknown length					
	16302	19585:	contig of 3284 bp in length					
			gap of unknown length					
	19586	23956:	contig of 4371 bp in length					
			gap of unknown length					
	23957	28296:	contig of 4340 bp in length					
			gap of unknown length					
	28297	31554:	contig of 3258 bp in length					
			gap of unknown length					
	31555	35178:	contig of 3624 bp in length					
			gap of unknown length					
	35179	39632:	contig of 4454 bp in length					
			gap of unknown length					
	39633	44084:	contig of 4452 bp in length					
			gap of unknown length					
	44085	48826:	contig of 4742 bp in length					
			gap of unknown length					
	48827	53193:	contig of 4367 bp in length					
			gap of unknown length					
	53194	56825:	contig of 3632 bp in length					
			gap of unknown length					


```
FEATURES
  source
    Location/Qualifiers
      1..145435
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="3p"
        /clone="RP11-439f4"

ORIGIN
  Query Match      3.4%; Score 76.4; DB 2; Length 145435;
  Best Local Similarity 69.3%; Pred. No. 4.5e-05;
  Matches 104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 119 CTTTTCCTGCGCAGTCTAGGTTGGTAAATGAGATGCGCTACTCTCTCGGACCATC 178
      |||
DB 79797 CTCCTGTCTCCCATGTCGCGATGTCGAGATGAGATGCTCTTCTATCTCGGACCATC 17856

QY 179 AAGAGGACATCATTTGACTATCACAGCTGACCTACGACCATGCTCAACCATCAGCA 238
      |||
DB 79857 CGCGGGATGTCGCTGCTGCTCTCTCTCCAGGATCAGATGATTAACCATCAGCC 79916

QY 239 AAGACAGCTCTCTTCCACCAAGGTGAGC 268
      |||
DB 79917 CGGCTGCTCTCTTACCAAGGTGGGC 79946

RESULT 10
AC022382/c
LOCUS Homo sapiens chromosome 3 clone RP11-266J6 map 3p, complete
DEFINITION AC022382
ACCESSION AC022382
VERSION AC022382.4 GI:24942870
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 185067)
  Bao, W., Bao, J., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
  Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,
  He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,
  Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,
  Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L.,
  Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,
  Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,
  Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X.,
  Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,
  Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
  Yu, J. and Yang, H.
  Chromosome 3p genomic sequence
  Unpublished
REFERENCE 2 (bases 1 to 185067)
  Wu, D., Hu, S., Dong, W., Zhang, X., Wang, J., Zhang, Y., Zhang, H.,
  Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y.,
  Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wang, X., Yu, B., Fan, H.,
  Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L.,
  Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L.,
  Feng, X., Yu, J. and Yang, H.
  Direct Submission
  Submitted (03-FEB-2000) Human Genomic Center, Institute of
  Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
  100101, P.R.China
REFERENCE 3 (bases 1 to 185067)
  Bao, W., Bao, J., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
  Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., He, L.,
  Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J.,
  Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W.,
  Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M.,
  Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, L.,
  Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F.,
  Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G.,
  Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, Y., Zhang, Y.,
  Zhang, X.,
```

15 unordered pieces.
AC011610
VERSION AC011610.11 GI:11055732
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 189430)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oman,F.R., Allen,C., Alebrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.K., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dunn-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homai,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulesged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,G., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwono,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189430)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 31, 2000 this sequence version replaced gi:9719552.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMNT
Center clone name: RP11-266J6
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159633 bases at least Q40
Consensus quality: 177583 bases at least Q30
Consensus quality: 184181 bases at least Q20

Estimated insert size: 183186; sum-of-contigs estimation
Estimated insert size: 198859; agarose-fp estimation
Quality coverage: 3.8x in Q20 bases; agarose-fp estimation
Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 40555: contig of 40555 bp in length
* 40556 40555: gap of unknown length
* 40556 83543: contig of 42888 bp in length
* 83544 83643: gap of unknown length
* 83644 99685: contig of 16042 bp in length
* 99686 99785: gap of unknown length
* 99786 119198: contig of 19413 bp in length
* 119199 119298: gap of unknown length
* 119299 134089: contig of 14791 bp in length
* 134090 134189: gap of unknown length
* 134190 149213: contig of 15024 bp in length
* 149214 159552: contig of 10239 bp in length
* 159553 159652: gap of unknown length
* 159653 168334: contig of 8682 bp in length
* 168335 168434: gap of unknown length
* 168435 175951: contig of 7517 bp in length
* 175952 176051: gap of unknown length
* 176052 179878: contig of 3827 bp in length
* 179879 179979: gap of unknown length
* 179980 184704: contig of 4726 bp in length
* 184705 184804: gap of unknown length
* 184805 185811: contig of 1007 bp in length
* 185812 185911: gap of unknown length
* 185912 186919: contig of 1008 bp in length
* 186920 187019: gap of unknown length
* 187020 188304: contig of 1285 bp in length
* 188305 188404: gap of unknown length
* 188405 189430: contig of 1026 bp in length.
FEATURES
source Location/Qualifiers
1..189430
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-266J6"
ORIGIN
Query Match 3.4%; Score 76.4; DB 2; Length 189430;
Best Local Similarity 69.3%; Pred. No. 4.3e-05;
Matches 104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 119 CTTTCTTTCTTGGCAGTCTAGTTGGTAAATAATGAGATGCGTACCTCTCTCGACCATC 178
Db 120335 CTCTGTCTCCCGAGTCCCGCATGGTCCAGATGAGATCCCTTCTTCTGACCACT 120276
Qy 179 AAGAGGGACATCATTTGACTATATCAGCCTTACAGCAGATGCTGAAACCACTACGCA 238
Db 120275 CGCGGGAGTGTGCTCGACTGTGCTTCTCTCCAGGATCAGATGATAAACCACTACGCC 120216
Qy 239 AAGACAGCTCTCTTACACCAAGGTGAGC 268
Db 120215 CGGGCTGGTCTCTTTACCAAAAGGTGGC 120186
RESULT 12
AC026196
LOCUS AC026196 191834 bp DNA linear HTG 29-MAY-2000

DEFINITION Homo sapiens chromosome 3 clone RP11-474F16 map 3p, WORKING DRAFT
ACCESSION AC026196
VERSION AC026196.3 GI:8101230
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Cai, T., Dong, X., Gao, Q., Gu, J., Gong, J., He, F., Kang, N., Lu, T., Ma, Q., Rong, L., Shen, Y., Tan, X., Wang, H., Xi, Y., Xu, Y., Yao, Z., Zheng, Z., Zhu, N., Zhou, Y., Qiang, B., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Yu, J. and Yang, H.
TITLE Chromosome 3p genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 191834)
AUTHORS Wang, R., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B., Qiao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Li, T., Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J. and Yang, H.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China
COMMENT On May 29, 2000 this sequence version replaced gi:8050894.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgsc@igtp.ac.cn
----- Project Information
Center project name:igtp project
Center clone name: RP11-474F16
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 5% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 185328 bases at least Q40
Consensus quality: 192037 bases at least Q30
Consensus quality: 196045 bases at least Q20
Insert size: 179116; sum-of-contigs
Quality coverage: 5.75x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 1632: contig of 1632 bp in length
* 1633 1732: gap of unknown length
* 1733 2953: contig of 1221 bp in length
* 2954 3053: gap of unknown length
* 3054 4934: contig of 1881 bp in length
* 4935 5034: gap of unknown length
* 5035 7618: contig of 2584 bp in length
* 7619 7718: gap of unknown length

7719 9805: contig of 2087 bp in length
9806 9905: gap of unknown length
9906 12859: contig of 2954 bp in length
12860 12959: gap of unknown length
12960 16605: contig of 3646 bp in length
16606 16705: gap of unknown length
16706 19419: contig of 2714 bp in length
19420 19519: gap of unknown length
19520 22470: contig of 2951 bp in length
22471 22570: gap of unknown length
22571 27404: contig of 4834 bp in length
27405 33396: contig of 5892 bp in length
33397 33496: gap of unknown length
33497 39063: contig of 5567 bp in length
39064 39163: gap of unknown length
39164 47186: contig of 8023 bp in length
47187 47287: gap of unknown length
47288 54906: contig of 7520 bp in length
54907 62077: gap of unknown length
62078 73474: contig of 11298 bp in length
73475 83952: contig of 10378 bp in length
83953 92882: contig of 15230 bp in length
92883 99383: gap of unknown length
99384 114261: contig of 14879 bp in length
114262 129675: contig of 15313 bp in length
129676 129774: gap of unknown length
129775 144971: contig of 15197 bp in length
144972 145071: gap of unknown length
145072 191834: contig of 46763 bp in length.
191835 191834: Location/Qualifiers
1. 191834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-474F16"
1. 1632
/feature="assembly_name:Contig17"
1733. 2953
/feature="assembly_name:Contig18"
3054. 4934
/feature="assembly_name:Contig20"
5035. 7618
/feature="assembly_name:Contig22"
7719. 9805
/feature="assembly_name:Contig23"
9906. 12859
/feature="assembly_name:Contig24"
12960. 16605
/feature="assembly_name:Contig25"
16706. 19419
/feature="assembly_name:Contig26"
19520. 22470
/feature="assembly_name:Contig27"
22571. 27404
/feature="assembly_name:Contig28"
27505. 33396
/feature="assembly_name:Contig29"
33497. 39063
/feature="assembly_name:Contig30"
39164. 47186
/feature="assembly_name:Contig31"
47287. 54806
/feature="assembly_name:Contig32"
54907. 62076
/feature="assembly_name:Contig33"

Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,G., Li,J., Li,L., Li,S., Liu,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L., Song,S., Sun,M., Sun,W., Sun,X., Tan,X., Tao,R., Wang,H., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Y., Zhang,L., Zhu,B., Zhu,N., Yu,J., and Yang,H.

Direct Submission
Submitted (02-JAN-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China

14 (bases 1 to 216911)
Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,X., Wang,Y., Wang,Z., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,Y., Zhang,Y., Zhang,L., Zhu,B., Zhu,N., Yu,J. and Yang,H.

Direct Submission
Submitted (25-MAR-2003) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China

On Mar 25, 2003 this sequence version replaced gi:12007686.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website: <http://hgsc.igtp.ac.cn>
<http://www.genomics.org.cn>
Contact:hgc@igtp.ac.cn
----- Project Information
Center project name:1# project
Center clone name: RP11-1020A11
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator: ET 55% of reads
Chemistry: Dye-terminator: Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 0 bases at least Q40
Consensus quality: 0 bases at least Q30
Consensus quality: 0 bases at least Q20
Insert size: 0; sum-of-contigs
Quality coverage: Infx in Q20 bases,sum-of-contigs

Location/Qualifiers
1. 216911
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-1020A11"

rch 3.4%; Score 76.4; DB 9; Length 216911;
l Similarity 69.3%; Pred. No. 4,2e-05;
104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

19 CTTTTCCTCCAGCTAGTGGTGAATAAATGAGATCCGTACCTCCTCTGGACCATC 178
30 CTTCCTGTCTCCCAAGTCCCGANGGTCCAGATGAGATCCCTTACTTCATCTGGACCAT 162971

779 AAGAGGGACATCATTTGACTATCAAGCTGACCTAGCACAGATGCTGAACCACTACGCA 238
70 CGCGGGATGTGCTGACTGTGCTTCTCTCCAGGATCAGATGATAAACCACTACGCC 162911

39 AAGACAGCCTCTTTCACCAACAAGGTGAGC 268

Db 162910 CGGGCTGGCTCTTTTACCACAAAGTGGGC 162881

RESULT 14

AC016947

LOCUS

DEFINITION Homo sapiens chromosome 3 clone RP11-532L22, linear HTG 19-AUG-2002
SEQUENCE, 18 unordered pieces.

ACCESSION

AC016947.24 GI:22203212

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 222994)

Murny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Albrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,

Honsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlsson,E., Kelly,S., King,U., King,L., Korvah,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,

Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapus,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,A., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

Rivas,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,

Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,

Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,

Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,

Williams,G., Williamson,A., Wlecyk,R., Wooden,S., Worley,K.,

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 222994)

Worley,K.C.

Direct Submission

Submitted (09-DEC-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 222994)

Worley,K.C.

Direct Submission

Submitted (19-AUG-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Aug 10, 2002 this sequence version replaced gi:21539052.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Drafting Center Code: BCM

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HMUI

Center clone name: RP11-532L22

----- Summary Statistics

Sequencing vector: Plasmid;

Sequencing vector: M13;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 215203 bases at least Q40

Consensus quality: 221146 bases at least Q30

Consensus quality: 225956 bases at least Q20

Estimated insert size: 192682; sum-of-contigs estimation

Estimated insert size: 213011; agarose-fp estimation

Quality coverage: 4x in Q20 bases; agarose-fp estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 18 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 2174: contig of 2174 bp in length

* 2175

* 2274: gap of unknown length

* 2275

* 4612: contig of 2338 bp in length

* 4613

* 4712: gap of unknown length

* 4713

* 6871: contig of 2159 bp in length

* 6872

* 6971: gap of unknown length

* 6972

* 9135: contig of 2164 bp in length

* 9136

* 9235: gap of unknown length

* 9236

* 11821: contig of 2586 bp in length

* 11822

* 11921: gap of unknown length

* 11922

* 14680: contig of 2759 bp in length

* 14681

* 14780: gap of unknown length

* 14781

* 17340: contig of 2560 bp in length

* 17341

* 17440: gap of unknown length

* 17441

* 20586: contig of 3146 bp in length

* 20587

* 20686: gap of unknown length

* 20687

* 2323: contig of 2637 bp in length

* 2324

* 23423: gap of unknown length

* 23424

* 26933: contig of 3510 bp in length

* 26934

* 27033: gap of unknown length

* 27034

* 27949: contig of 2716 bp in length

* 27950

* 29849: gap of unknown length

* 29850

* 32693: contig of 2844 bp in length

* 32694

* 32793: gap of unknown length

* 32794

* 35852: contig of 3059 bp in length

* 35853

* 35952: gap of unknown length

* 35953

* 38534: gap of unknown length

* 38535

* 38635: gap of unknown length

* 42235

* 42236: gap of unknown length

* 42335

* 42336: contig of 3362 bp in length

* 45697

* 45796: gap of unknown length

* 45977

* 51396: contig of 5600 bp in length

* 51397

* 51496: gap of unknown length

* 51497

* 222994: contig of 171498 bp in length.

Location/Qualifiers

1. 222994

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="3"

/clone="RP11-532L22"

FEATURES

source

ORIGIN

```
Query Match          3.4%; Score 76.4; DB 2; Length 222994;
Best Local Similarity 69.3%; Pred. No. 4.2e-05;
Matches 104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 119 CTTTTCCTTCCTCAGTGTAGGTTGGTAAAAAATGAGATGCGTACCTCTCTGGACCATC 178
Db 13233 CTCCTGTCCTCCAGTCCGCGATGGTCCAGATGAGATCCCTACTTCACTGGACCACT 13292

Qy 179 AAGAGGACATCATTTGACTATCAGCTGACCTACCTACGACCATGCTGAACCACTAGCCA 238
Db 13293 CGGCGGATGTGCTCGAGTGTGCTTCTCTTCCAAGGATCAGATGATTAACCACTAGGCC 13352

Qy 239 AAGACAGCCTCTTCCACCAACCAAGGTGAGC 268
Db 13353 CGGCTGGCTCTTACCAACAAGGTGGC 13382

RESULT 15
AC068315 155313 bp DNA linear HTG 27-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-224P21 map 3p, WORKING DRAFT
DEFINITION AC068315
AC068315.4 GI:10312236
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155313)
AUTHORS Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, J., Li, J., Li, L.,
Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y.,
Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y.,
Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R.,
Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
Yang, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H.,
Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,
Zhang, Z., Zhu, B., Yu, J. and Yang, H.
Unpublished
TITLE Chromosome 3p genomic sequence
JOURNAL
REFERENCE 2 (bases 1 to 155313)
AUTHORS Bao, W., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
Li, G., Li, C., Bao, Q., Bao, Q., Bao, J., Wang, X., Song, L., Guo, D.,
Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.
and Yang, H.
Direct Submission
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Sep 27, 2000 this sequence version replaced gi:8101156.
-----Genome Center
Center: Beijing Center
Center code: Beijing
Website: http://hg.igtp.ac.cn
http://www.genomics.org.cn
Contact: hg@igtp.ac.cn
-----Project Information
Center project name: 1k project
Center clone name: RP11-224P21
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 154843 bases at least Q40
Consensus quality: 165011 bases at least Q30
Consensus quality: 168985 bases at least Q20
Insert size: 139544; sum-of-contigs
```

```
Quality coverage: 5.36x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1139: contig of 1139 bp in length
* 1140 1239: gap of unknown length
* 1240 2448: contig of 1209 bp in length
* 2449 2548: gap of unknown length
* 2549 3935: contig of 1387 bp in length
* 3936 4035: gap of unknown length
* 4036 6131: contig of 2096 bp in length
* 6132 6231: gap of unknown length
* 6232 7429: contig of 1198 bp in length
* 7430 7529: gap of unknown length
* 7530 9054: contig of 1525 bp in length
* 9055 9154: gap of unknown length
* 9155 10950: contig of 1796 bp in length
* 10951 11050: gap of unknown length
* 11051 12637: contig of 1587 bp in length
* 12638 12737: gap of unknown length
* 12738 15121: contig of 2384 bp in length
* 15122 15221: gap of unknown length
* 15222 17220: contig of 1999 bp in length
* 17221 17320: gap of unknown length
* 17321 20436: contig of 3116 bp in length
* 20437 20536: gap of unknown length
* 20537 23761: contig of 3225 bp in length
* 23762 23861: gap of unknown length
* 23862 26190: contig of 2329 bp in length
* 26191 26290: gap of unknown length
* 26291 28871: contig of 2581 bp in length
* 28872 28971: gap of unknown length
* 28972 33062: contig of 4091 bp in length
* 33063 33162: gap of unknown length
* 33163 37033: contig of 3871 bp in length
* 37034 37133: gap of unknown length
* 37134 43465: contig of 6332 bp in length
* 43466 43565: gap of unknown length
* 43566 48805: contig of 5240 bp in length
* 48806 48905: gap of unknown length
* 48906 56158: contig of 7253 bp in length
* 56159 56258: gap of unknown length
* 56259 63410: contig of 7151 bp in length
* 63410 63509: gap of unknown length
* 63510 72539: contig of 9030 bp in length
* 72540 72639: gap of unknown length
* 72640 83389: contig of 10750 bp in length
* 83390 83489: gap of unknown length
* 83490 91426: contig of 7937 bp in length
* 91427 91526: gap of unknown length
* 91527 99509: contig of 7983 bp in length
* 99510 99609: gap of unknown length
* 99610 108238: contig of 8629 bp in length
* 108239 108338: gap of unknown length
* 108339 119079: contig of 10741 bp in length
* 119080 119179: gap of unknown length
* 119180 135925: contig of 16746 bp in length
* 135926 136025: gap of unknown length
* 136026 155313: contig of 19288 bp in length.
*
* Location/Qualifiers
* 1. 155313
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="3"
* /map="3p"
* /clone="RP11-224P21"
```

FEATURES
source

```
misc_feature 1. .1139
misc_feature /note="assembly_name:Contig32"
misc_feature 1240. .2448
misc_feature /note="assembly_name:Contig33"
misc_feature 2549. .3935
misc_feature /note="assembly_name:Contig34"
misc_feature 4036. .6131
misc_feature /note="assembly_name:Contig35"
misc_feature 6232. .7429
misc_feature /note="assembly_name:Contig36"
misc_feature 7530. .9054
misc_feature /note="assembly_name:Contig37"
misc_feature 9155. .10950
misc_feature /note="assembly_name:Contig38"
misc_feature 11051. .12837
misc_feature /note="assembly_name:Contig39"
misc_feature 12738. .15121
misc_feature /note="assembly_name:Contig40"
misc_feature 15222. .17220
misc_feature /note="assembly_name:Contig41"
misc_feature 17321. .20436
misc_feature /note="assembly_name:Contig42"
misc_feature 20537. .23761
misc_feature /note="assembly_name:Contig43"
misc_feature 23862. .26190
misc_feature /note="assembly_name:Contig44"
misc_feature 26291. .28871
misc_feature /note="assembly_name:Contig45"
misc_feature 28972. .33062
misc_feature /note="assembly_name:Contig46"
misc_feature 33163. .37033
misc_feature /note="assembly_name:Contig47"
misc_feature 37134. .43465
misc_feature /note="assembly_name:Contig48"
misc_feature 43566. .48805
misc_feature /note="assembly_name:Contig49"
misc_feature 48906. .56158
misc_feature /note="assembly_name:Contig50"
misc_feature 56259. .63409
misc_feature /note="assembly_name:Contig51"
misc_feature 63510. .72539
misc_feature /note="assembly_name:Contig52
misc_feature clone_end:SP6
misc_feature vector_side:right"
misc_feature 72640. .83389
misc_feature /note="assembly_name:Contig53"
misc_feature 83490. .91426
misc_feature /note="assembly_name:Contig54"
misc_feature 91527. .99509
misc_feature /note="assembly_name:Contig55"
misc_feature 99610. .108238
misc_feature /note="assembly_name:Contig56"
misc_feature 108339. .119079
misc_feature /note="assembly_name:Contig57"
misc_feature 119180. .135925
misc_feature /note="assembly_name:Contig58"
misc_feature 136026. .155313
misc_feature /note="assembly_name:Contig59"
```

ORIGIN

```
Query Match 3.3%; Score 73.2; DB 2; Length 155313;
Best Local Similarity 68.0%; Pred.No.0.00019;
Matches 102; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 119 CTTTTTCTTGCCAGCTAGGTTGGTAAATAATGAGATGCCGTACTCTCTGGACCATC 178
Db |||||
QY 48631 CTCCTGTCTCCCGAGTCCGGATGTCAGATGATGCCCTACTTCTATCTGGACCATC 48690
Db |||||

QY 179 AAGAGGGACATCATTTGACTATCACAGCCTGACCTACGACGATGCTGAACCACTACGCA 238
Db |||||
QY 48691 CGCGGGATGTGTCGACTGACGCTTCTCTCCAAGGATGATGATTAACCACTACCCC 48750
Db |||||
QY 239 AAGACAGCCTCTCTTACCACCAAGGTGAGC 268
```

```
Db 48751 CGGGCTGGCTCCTTTACCAAGGTGGGC 48780
|||||
```

Search completed: September 25, 2005, 11:27:01
Job time : 10407.7 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 01:51:04 ; Search time 1359.73 Seconds
(without alignment)
9756.464 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgctgcagacagcccgccg.....gcctccacagggccctccc 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_l6Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2241	100.0	2241	12 ADJ93378	Adj93378 Human BGS
2	2241	100.0	101270	12 ADQ17814	Adq17814 Human sof
3	1795.8	80.1	3465	12 ADJ93363	Adj93363 Human BGS
4	1742.8	77.8	3554	12 ADJ93364	Adj93364 Human tub
5	272.4	12.2	1838	12 ADJ93357	Adj93357 Human BGS
6	272.4	12.2	1939	12 ADJ93361	Adj93361 Human BGS
7	138.4	6.2	1859	12 ADJ93362	Adj93362 Human BGS
8	71.6	3.2	5282	12 ADQ64426	Adq64426 Novel hum
9	69.2	3.1	452	6 ABQ97805	Abq97805 Mouse ES
10	67.8	3.0	1042	5 AAS72428	Aas72428 DNA encod
11	67.8	3.0	2250	5 AAS69383	Aas69383 DNA encod
12	60.8	2.7	5452	10 ADC86736	Adc86736 Human GPC
13	58.2	2.6	629	13 ACN54594	Acn54594 Cotton an
14	58	2.6	1337	2 AAZ17263	Aaz17263 Human gen
15	57.4	2.6	588	13 ACN54596	Acn54596 Cotton an
16	57	2.5	1327	6 ABQ68452	Abq68452 Listeria
17	56.8	2.5	3163	10 ADC87060	Adc87060 Human GPC
18	56	2.5	12733	6 ABK98631	Abk98631 Vector pE
19	56	2.5	12733	9 ACD13882	Acld13882 L. lactis
20	56	2.5	12739	6 ABK98592	Abk98592 Vector pE

21	56	2.5	12739	9 ACD13843	Acld13843 Plasmid p
22	55.6	2.5	1000	3 AAA02484	Aaa02484 Human col
23	55.2	2.5	2000	8 ADA71938	Ada71938 Rice gene
24	55	2.5	3133	10 ADC86738	Adc86738 Human GPC
25	55	2.5	28198	10 ADG37080	Adg37080 Mouse pla
26	54.4	2.4	110000	12 ADQ97050_0	Adq97050 Human can
27	54.4	2.4	114955	2 AAX53491	Aax53491 Human ade
28	54.2	2.4	931	11 ACN87435	Acn87435 Breast ca
29	53.8	2.4	600	6 ABQ52497	Abq52497 Oligonuc1
30	53.8	2.4	600	6 ABQ52496	Abq52496 Oligonuc1
31	53.6	2.4	2188	2 AAZ77506	Aaz77506 Human ova
32	53.2	2.4	1065	6 ABT09682	Abt09682 Human PAL
33	53.2	2.4	40491	11 ACN44798	Acn44798 Human gen
34	52.8	2.4	922	10 ADC86708	Adc86708 Human GPC
35	52.8	2.4	3198	2 AAX02974	Aax02974 Human lu-
36	52.2	2.3	318	3 AAA38184	Aaa38184 Primer us
37	52.2	2.3	778	4 AAH08565	Aah08565 Human cDN
38	52.2	2.3	2326	4 AAH16735	Aah16735 Human cDN
39	51.2	2.3	320	3 AAA38185	Aaa38185 Primer us
40	51.2	2.3	1117	10 ADC86688	Adc86688 Human GPC
41	51	2.3	2561	5 AAH28500	Aah28500 Rabbit lo
42	50.6	2.3	500	12 ADP85918	Adp85918 Synthetic
43	50.6	2.3	1218	3 AAA02488	Aaa02488 Human col
44	50	2.2	320	3 AAA38186	Aaa38186 Primer us
45	49.8	2.2	420	13 ACN51217	Acn51217 Cotton an

ALIGNMENTS

RESULT 1

ADJ93378

ID ADJ93378 standard; DNA; 2241 BP.

XX ADJ93378;

XX AC ADJ93378;

XX DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 gene promoter DNA sequence.

XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nocotropic; antiparkinsonian; antiarthritic; antiseborrheic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.

XX OS Homo sapiens.

XX XX WO2004005487-A2.

XX XX 15-JAN-2004.

XX XX 09-JUL-2003; 2003WO-US021605.

XX XX 09-JUL-2002; 2002US-0394725P.

XX XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX XX Feder JN, Wu S, Nelson TC;

XX XX WPI; 2004-099381/10.

XX XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, e.g.
PT useful for preventing, treating or ameliorating a medical condition, e.g.

aberrant cellular proliferation, reproductive disorders or testicular disorders.

Claim 1; SEQ ID NO 27; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase-like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cytostatic, respiratory

-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, anti-inflammatory, anabolic, hypertensive, osteopathic, nootropic, antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial, immunosuppressive, antiseborrheic or dermatological activity acting as tyrosine ligase modulators. In addition, the disclosed sequences may be useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a gene which encodes the human BGS-42 protein of the invention. The present sequence is that of the promoter of the gene which encodes the human BGS-42 sequence of the invention.

Sequence 2241 BP; 388 A; 598 C; 769 G; 386 T; 0 U; 0 Other;

Query Match 100.0%; Score 2241; DB 12; Length 2241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCCGAGAGAGCCGGCTGGGATCTCACGGTGCACGGCCATTTCTCTGTATGGTT 60
Db 1 TGCCGAGAGAGCCGGCTGGGATCTCACGGTGCACGGCCATTTCTCTGTATGGTT 60
Qy 61 GCCTCACTCCATCCATCTGGCTGGTGTGAACCTTGAATCTGGGCATCAATAAGACT 120
Db 61 GCCTCACTCCATCCATCTGGCTGGTGTGAACCTTGAATCTGGGCATCAATAAGACT 120
Qy 121 TTTTCTTGCCAGTCTAGGTTGGTAAATAATGAGATCCGTPACTCTCTGGACCATCAA 180
Db 121 TTTTCTTGCCAGTCTAGGTTGGTAAATAATGAGATCCGTPACTCTCTGGACCATCAA 180
Qy 181 GAGGGACATCACTATCAAGCCTGACCTAGACAGCATGCTGAACCACTACGCAAA 240
Db 181 GAGGGACATCACTATCAAGCCTGACCTAGACAGCATGCTGAACCACTACGCAAA 240
Qy 241 GACAGCCTCTTACACCAAGGTGAGCCGCGCGCTTGAGCGTGGCGGAGCCGGCAA 300
Db 241 GACAGCCTCTTACACCAAGGTGAGCCGCGCGCTTGAGCGTGGCGGAGCCGGCAA 300
Qy 301 GGGTGGTGGGAGGAGCTTCTGACGCCATTAAGGACCCCTCGGTGGCTGGTCAAGTGGCCAC 360
Db 301 GGGTGGTGGGAGGAGCTTCTGACGCCATTAAGGACCCCTCGGTGGCTGGTCAAGTGGCCAC 360
Qy 361 CAGTCACCTCTGGTGCATCCAGGACTCTGTGCTTAAGCCGTCGCCAGAACTCACTCG 420
Db 361 CAGTCACCTCTGGTGCATCCAGGACTCTGTGCTTAAGCCGTCGCCAGAACTCACTCG 420
Qy 421 GTGCCACCCACCCAGCACCCCTGTGCCCTTTGTCTGTGTCTCTGGTGAATCCGG 480
Db 421 GTGCCACCCACCCAGCACCCCTGTGCCCTTTGTCTGTGTCTCTGGTGAATCCGG 480
Qy 481 GGCCCCAGAGCTCCCTCTCAGGGCACAAGAGATGGGGGCTGACTGGGGGCTGC 540

Db 481 GGCCCCAGAGCTCCCTCTCAGGGCACAAGATGGGGCTGACTGGGGGCTGC 540
Qy 541 CACCGGGCTTTGGTGTCTGAGGGGGCTGTGGGACCCCGAGGGGAAGAGGTGCCCGCTCCCCC 600
Db 541 CACCGGGCTTTGGTGTCTGAGGGGGCTGTGGGACCCCGAGGGGAAGAGGTGCCCGCTCCCCC 600
Qy 601 CAGCACCGGGCACTCCCCCGCCTCCCGCGCTCCAGGCCCTCGACAAAGCAGCTTGGCAC 660
Db 601 CAGCACCGGGCACTCCCCCGCCTCCCGCGCTCCAGGCCCTCGACAAAGCAGCTTGGCAC 660
Qy 661 ACCCTTCCACCAAGGGCTCAGGGCTTCCCGAGAGAGGGTGTGGGGACCCCTG 720
Db 661 ACCCTTCCACCAAGGGCTCAGGGCTTCCCGAGAGAGGGTGTGGGGACCCCTG 720
Qy 721 CTGAGGAGGGGTCCCGGAAGAGGGCTTCCCTGGCACAGAGTCCCTCTCAGCAGGCGCA 780
Db 721 CTGAGGAGGGGTCCCGGAAGAGGGCTTCCCTGGCACAGAGTCCCTCTCAGCAGGCGCA 780
Qy 781 GGTGCGGCTGCTCAGCACAGTGTGGGGCGAGGTGCAGCAAGGTTCCTTCCCGCAC 840
Db 781 GGTGCGGCTGCTCAGCACAGTGTGGGGCGAGGTGCAGCAAGGTTCCTTCCCGCAC 840
Qy 841 TAATACCCAGGGTCAAGCCAGCCAGCCGCTGCTAGTGGCAACATGGCCCTTCAAGA 900
Db 841 TAATACCCAGGGTCAAGCCAGCCAGCCGCTGCTAGTGGCAACATGGCCCTTCAAGA 900
Qy 901 CCCGCTGTGCAGAGCCAGTCAAGCGTGGCTGTGTTTCTGGGGCCAGCCACTTTTTC 960
Db 901 CCCGCTGTGCAGAGCCAGTCAAGCGTGGCTGTGTTTCTGGGGCCAGCCACTTTTTC 960
Qy 961 TTAACAGGGTGAACAACAGACATCGGGGTGCGGGGACTTCAGGATATGCTCTTAATG 1020
Db 961 TTAACAGGGTGAACAACAGACATCGGGGTGCGGGGACTTCAGGATATGCTCTTAATG 1020
Qy 1021 GCCTGTCTACTCTCGGTGTGGTGGGAGTGTGGGATGTGGTGGGCGCCCTG 1080
Db 1021 GCCTGTCTACTCTCGGTGTGGTGGGAGTGTGGGATGTGGTGGGCGCCCTG 1080
Qy 1081 CAGCGCTCTCTCTCTCTGGGTGTGAGGTGTGGGACACCTGTGGCCCGCTTTGGGG 1140
Db 1081 CAGCGCTCTCTCTCTCTGGGTGTGAGGTGTGGGACACCTGTGGCCCGCTTTGGGG 1140
Qy 1141 CCTGGAGGAGCCCAAGTGCACCCAGCCCTCGGTGGCTCAGCTGACAGCTCCCATTC 1200
Db 1141 CCTGGAGGAGCCCAAGTGCACCCAGCCCTCGGTGGCTCAGCTGACAGCTCCCATTC 1200
Qy 1201 CTGGGGCTGCACAGCCCTTCACTGCTCAAGGATGGTCTGAGCCTGGGCTGCACATCC 1260
Db 1201 CTGGGGCTGCACAGCCCTTCACTGCTCAAGGATGGTCTGAGCCTGGGCTGCACATCC 1260
Qy 1261 ATGCTCCACCGGTGGAGGGGTGTCTGGGGCCCTGGGGGACCGTATATCTGGGGAGG 1320
Db 1261 ATGCTCCACCGGTGGAGGGGTGTCTGGGGCCCTGGGGGACCGTATATCTGGGGAGG 1320
Qy 1321 GGGAGGGATCCGCGCTAGAGGGTGTGTGGTGTCCCATAGCCAGGCAAGTGTGAGTGG 1380
Db 1321 GGGAGGGATCCGCGCTAGAGGGTGTGTGGTGTCCCATAGCCAGGCAAGTGTGAGTGG 1380
Qy 1381 ACGGCCGCTCTCGCTTGTGAGGAGGAGAGTGAAGATGGCTGAGCTGCCCTCGGCTAA 1440
Db 1381 ACGGCCGCTCTCGCTTGTGAGGAGGAGAGTGAAGATGGCTGAGCTGCCCTCGGCTAA 1440
Qy 1441 CAGCCACTGTCTCTGGAGATCGGGTGTGGTGAACATGCGGAGCTGCCCTGGTACGT 1500
Db 1441 CAGCCACTGTCTCTGGAGATCGGGTGTGGTGAACATGCGGAGCTGCCCTGGTACGT 1500
Qy 1501 CCCGGCAACCCCGACTCTTTCACAGCTGTCTGCAAGCTGTGACCGAGAGTGAACA 1560
Db 1501 CCCGGCAACCCCGACTCTTTCACAGCTGTCTGCAAGCTGTGACCGAGAGTGAACA 1560
Qy 1561 GCAGGATTCCTGGTGAAGTGAAGAGAGCGAGGGGCCAGTGTGTCAGCAGAGAGG 1620

QY 661 ACCCTTCCCAACCAAGGCCAGGGCTGAGGCTGCTGCCAGGACGACGAGGTGTGGGACCCCTG 720
DB 59478 ACCCTTCCCAACCAAGGCCAGGGCTGAGGCTGCTGCCAGGACGACGAGGTGTGGGACCCCTG 59419
QY 721 CTGAGGAGGGGTCCCGAAGAGAGGGGCTTCCCTGGCAAGAGGTCCCTCTCAGCAGGCCA 780
DB 59418 CTGAGGAGGGGTCCCGAAGAGAGGGGCTTCCCTGGCAAGAGGTCCCTCTCAGCAGGCCA 59359
QY 781 GGTCCGGCTGCTCAGCACAGTGTGGGGCGAGGTGCAGGACAGGTTCCTTCCGACAC 840
DB 59358 GGTCCGGCTGCTCAGCACAGTGTGGGGCGAGGTGCAGGACAGGTTCCTTCCGACAC 59299
QY 841 TAATACCCCAAGGTCAGGCCAGCCAGCCAGCTGTCTAGTGGCAACATGGCCCCCTTCAAAGA 900
DB 59298 TAATACCCCAAGGTCAGGCCAGCCAGCCAGCTGTCTAGTGGCAACATGGCCCCCTTCAAAGA 59239
QY 901 CCCGCTGTGAGAGCCAGTCAAGCTGCGCTGTGTTTTCTGGGGGCCAGCACATTTTTTTC 960
DB 59238 CCCGCTGTGAGAGCCAGTCAAGCTGCGCTGTGTTTTCTGGGGGCCAGCACATTTTTTTC 59179
QY 961 TTAACAGGTTGACAAAACAGACATCGGGGTGCGGGACTTTCAGATATGCTCTCTAATG 1020
DB 59178 TTAACAGGTTGACAAAACAGACATCGGGGTGCGGGACTTTCAGATATGCTCTCTAATG 59119
QY 1021 GCGCGTGTCTACCTCGCTGTGGGTCAGTGTGGGCAATGTGGCTGCTGCCGCCCTG 1080
DB 59118 GCGCGTGTCTACCTCGCTGTGGGTCAGTGTGGGCAATGTGGCTGCTGCCGCCCTG 59059
QY 1081 CAGCCGCTCTCTCTCTCTGGGTTCTGAGGTGTGGACACCTGTGGCCGCTTGGGG 1140
DB 59058 CAGCCGCTCTCTCTCTCTGGGTTCTGAGGTGTGGACACCTGTGGCCGCTTGGGG 58999
QY 1141 CTTGGAGGAGCCCACTGACCCAGCCGCTCGGTGCGCTCAGCTGCAGCCTCCCATTC 1200
DB 58998 CTTGGAGGAGCCCACTGACCCAGCCGCTCGGTGCGCTCAGCTGCAGCCTCCCATTC 58939
QY 1201 CTGGGGCTTGCCAGCCCTCTCACTGCTCAAGATGCTGTGAGCCTTGGCTTGCACATCC 1260
DB 58938 CTGGGGCTTGCCAGCCCTCTCACTGCTCAAGATGCTGTGAGCCTTGGCTTGCACATCC 58879
QY 1261 ATGCTCCACCGGTGAGGGGTGTCTGGGGCTTGGGGACCGGTATATCTGGGGAGG 1320
DB 58878 ATGCTCCACCGGTGAGGGGTGTCTGGGGCTTGGGGACCGGTATATCTGGGGAGG 58819
QY 1321 GGGAGGATCCGCCCTAGGGGTGTGTGTGCCCATAGCCAGGCAAGTCTGAGGTGG 1380
DB 58818 GGGAGGATCCGCCCTAGGGGTGTGTGTGCCCATAGCCAGGCAAGTCTGAGGTGG 58759
QY 1381 ACGCCCTGCTCTCGCTTGTGAGCGAGCGACAGCTGAGCTGAGCTGCCCTCTCTGCCCTAA 1440
DB 58758 ACGCCCTGCTCTCGCTTGTGAGCGAGCGACAGCTGAGCTGAGCTGCCCTCTCTGCCCTAA 58699
QY 1441 CAGCCACTGTCTTGAGATCGGGCTGTGCGTGAACATCGGAGCGCTGCCCTGGTACGT 1500
DB 58698 CAGCCACTGTCTTGAGATCGGGCTGTGCGTGAACATCGGAGCGCTGCCCTGGTACGT 58639
QY 1501 CCGGGCAACCCGACTCTTCTCCACCGCTGTACAGCTCTGCACCGAGAGTGAGCA 1560
DB 58638 CCGGGCAACCCGACTCTTCTTCCACCGCTGTCTACAGCTCTTGCACCGAGAGTGAGCA 58579
QY 1561 GCAGGAGTCTCTGGGTAACTGAGGAGACGGCAGAGGGCCCCAGTGTCTCAGCAGAGAGG 1620
DB 58578 GCAGGAGTCTCTGGGTAACTGAGGAGACGGCAGAGGGCCCCAGTGTCTCAGCAGAGAGG 58519
QY 1621 CTTCTAGAAAGATCCCTCTGGTGTGAGACAGACTGTATGGGGCAGGGTCTTGAGATAGAG 1680
DB 58518 CTTCTAGAAAGATCCCTCTGGTGTGAGACAGACTGTATGGGGCAGGGTCTTGAGATAGAG 58459
QY 1681 GACCGGGAGAGGCTCCCATGGTTCATGGTCATGGCAGTACAGAGGCCAGGGGCCCGGG 1740
DB 58458 GACCGGGAGAGGCTCCCATGGTTCATGGTTCATGGCAGTACAGAGGCCAGGGGCCCGGG 58399

QY 1741 AGGAGGGGGGCGAGTCAAGAGCTGTGTGTTCTGTAGGCTTGTGGAACGCGCGCG 1800
DB 58398 AGGAGGGGGGCGAGTCAAGAGCTGTGTGTTCTGTAGGCTTGTGGAACGCGCGCG 58339
QY 1801 AGCAGCTGGGGCAAGCGCGCTCCGTGTCGAGGCGGAGGGGTGCACTGGAGCGCCG 1860
DB 58338 AGCAGCTGGGGCAAGCGCGCTCCGTGTCGAGGCGGAGGGGTGCACTGGAGCGCCG 58279
QY 1861 AGTTCACAGAGACTCTCAGGGAGAAAGCGAGGCGAGACTTGGGGGTGTGGAGCAGGCTCG 1920
DB 58278 AGTTCACAGAGACTCTCAGGGAGAAAGCGAGGCGAGACTTGGGGGTGTGGAGCAGGCTCG 58219
QY 1921 GCCCATCGCTATAGGGCGGGAAGTGGGGCGGGTTGGGGAGCCTCCGTGGCCCTGGCC 1980
DB 58218 GCCCATCGCTATAGGGCGGGAAGTGGGGCGGGTTGGGGAGCCTCCGTGGCCCTGGCC 58159
QY 1981 CCACTCTCAGTCCCGCAGTCTGCTCGGACAGATAGGGCGAGGCTGTGCTCTTTTCAG 2040
DB 58158 CCACTCTCAGTCCCGCAGTCTGCTCGGACAGATAGGGCGAGGCTGTGCTCTTTTCAG 58099
QY 2041 AAGACTTCCGGCGCACCATGGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACAGAGCT 2100
DB 58098 AAGACTTCCGGCGCACCATGGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACAGAGCT 58039
QY 2101 GCAGCAGGACGACGACAAAGCCAGGAGACAGAGGAGGAGCGCGGAGCAGCGACC 2160
DB 58038 GCAGCAGGAGCAGCAGAAAGCAAGCCAGGAGACAGAGGAGGAGCGCGGAGCAGCGACC 57979
QY 2161 TGAGCAGCAGGCAAGGTGCGCTGGGCGGGGAGGAGCAGTGCAGGCCACAGAGCTCGG 2220
DB 57978 TGAGCAGCAGGCAAGGTGCGCTGGGCGGGGAGGAGCAGTGCAGGCCACAGAGCTCGG 57919
QY 2221 GGCTTCCACAGGGGGCTTCCC 2241
DB 57918 GGCTTCCACAGGGGGCTTCCC 57898
RESULT 3
ADJ93363
ID ADJ93363 standard; DNA; 3465 BP.
AC ADJ93363;
XX
DT 06-MAY-2004 (first entry)
XX Human BGS-42 protein-related DNA clone C SeqID11.
DE testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
OS Homo sapiens.
XX
XX WO2004005487-A2.
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;
PI WPI; 2004-099381/10.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
PS Example 4; SEQ ID NO 11; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -gen, gastrointestinal-gen, neuroprotective, endocrine-gen,
CC antinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antithrombotic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiepileptic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match
Best Local Similarity 80.1%; Score 1795.8; DB 12; Length 3465;
Matches 1797; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 381 CCCAGGACTCTGTGCTTAAGGCGGTGGCCAGAAATCACTCGGTGCCACCCACCCACCCAGC 440
DB 1 CCCAGGACTCTGTGCTTAAGGCGGTGGCCAGAAATCACTCGGTGCCACCCACCCACCCAGC 60
QY 441 ACCCTGTGCCCTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 500
DB 61 ACCCTGTGCCCTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 120
QY 501 CAGGCGACAGAGCCCAAGATGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGA 560
DB 121 CAGGCGACAGAGCCCAAGATGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGA 180
QY 561 GGGGGCTGTGGGACCCCGAGGGAAGAGTGGCGCTCCCGCCAGCAGCCGGGCACTCCCGCG 620
DB 181 GGGGGCTGTGGGACCCCGAGGGAAGAGTGGCGCTCCCGCCAGCAGCCGGGCACTCCCGCG 240
QY 621 CTTCCCGCGGCTTCAGCGCTTGCACAAAGAGCTTGGCCACACCTTCCCGCCAGAGGCCCA 680
DB 241 CTTCCCGCGGCTTCAGCGCTTGCACAAAGAGCTTGGCCACACCTTCCCGCCAGAGGCCCA 300
QY 681 GGGCTGAGGCTTCCCGAGGACCCAGGCTGGGGACCTGCTGAGGGAGGGGTCCCGGAA 740
DB 301 GGGCTGAGGCTTCCCGAGGACCCAGGCTGGGGACCTGCTGAGGGAGGGGTCCCGGAA 360
QY 741 GAGGGGCTTCCCTGGCAGAGGTCCTCTCAGCAGGCGCAGGTGGCGGTGCTCAGCACA 800
DB 361 GAGGGGCTTCCCTGGCAGAGGTCCTCTCAGCAGGCGCAGGTGGCGGTGCTCAGCACA 420

QY 801 GTGTGGGGCGGAGGTGTCAGGACAAGTTTCCTTCGGCACCTTAATACCCAGGGTCAGGCC 860
DB 421 GTGTGGGGCGGAGGTGTCAGGACAAGTTTCCTTCGGCACCTTAATACCCAGGGTCAGGCC 480
QY 861 AGCGCCAGCGTGTCTAGTGGCAACATAGCCGCCCTTCAAAAGACCCCGTGTGCAGAGCCAGTC 920
DB 481 AGCGCCAGCGTGTCTAGTGGCAACATAGCCGCCCTTCAAAAGACCCCGTGTGCAGAGCCAGTC 540
QY 921 AGCGTGCCTGTGTCTTCTGGGGGCGCAGCCACTTTTCTTTTCTTTTAAAGGGGTGACAAACAG 980
DB 541 AGCGTGCCTGTGTCTTCTGGGGGCGCAGCCACTTTTCTTTTCTTTTAAAGGGGTGACAAACAG 600
QY 981 ACATCGGGGTCGGGGGACTTCACGATATGCTCTCTAATGCGCGTGTACCTCGCCTGT 1040
DB 601 ACATCGGGGTCGGGGGACTTCACGATATGCTCTCTAATGCGCGTGTACCTCGCCTGT 660
QY 1041 GGTGGGCAAGTGTGGGGCAATGTGGCTTCCGGGCGCTTCGAGCGGCTCTCTCTCTCTCT 1100
DB 661 GGTGGGCAAGTGTGGGGCAATGTGGCTTCCGGGCGCTTCGAGCGGCTCTCTCTCTCTCTCT 720
QY 1101 GGGGTCTGAGGTGTGGGACACCTGTGGCCCGTTTGGGGGCTTGGAGGAGGCGCCAGTGC 1160
DB 721 GGGGTCTGAGGTGTGGGACACCTGTGGCCCGTTTGGGGGCTTGGAGGAGGCGCCAGTGC 780
QY 1161 CCACGCGCTCGGTGCGCTCAGCTGCAGCTCCCATTTCCCTGGGGCGCTGCCACGCCCTC 1220
DB 781 CCACGCGCTCGGTGCGCTCAGCTGCAGCTCCCATTTCCCTGGGGCGCTGCCACGCCCTC 840
QY 1221 CACTGCTCTCAAGGATGCGCTCTGAGCTTGGCTGACATCCATGCTCCACCGGCTGAGG 1280
DB 841 CACTGCTCTCAAGGATGCGCTCTGAGCTTGGCTGACATCCATGCTCCACCGGCTGAGG 900
QY 1281 GGGTGTCTTGGGGCGCTGGGGGACCGTATATCTGGGGAGGGGAGGATTCGGCCCTTAGGG 1340
DB 901 GGGTGTCTTGGGGCGCTGGGGGACCGTATATCTGGGGAGGGGAGGATTCGGCCCTTAGGG 960
QY 1341 GTGCTGTGTGTGTCCTTACCCAGGCAAGTCTCAGGTGGAGCGGCCCTCTCGCCTTGT 1400
DB 961 GTGCTGTGTGTGTCCTTACCCAGGCAAGTCTCAGGTGGAGCGGCCCTCTCGCCTTGT 1020
QY 1401 GAGCGAGGCGAGCTGAGCTGAGCTGCGCCCTCTGCTCTTAAACACGCTCTCTCTGGAGA 1460
DB 1021 GAGCGAGGCGAGCTGAGCTGAGCTGCGCCCTCTGCTCTTAAACACGCTCTCTCTGGAGA 1080
QY 1461 TCGGGCTGTGCTGAAACATCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1520
DB 1081 TCGGGCTGTGCTGAAACATCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1521 TCTTTCCACCGCTTACAGCTCTGCAACCGAGGTGAGCAGGAGTTCCTGGGTAAAGT 1580
DB 1141 TCTTTCCACCGCTTACAGCTCTGCAACCGAGGTGAGCAGGAGTTCCTGGGTAAAGT 1200
QY 1581 GAGGAGCGGACAGAGGGGCGCCAGTGTCTCAGCAGAGAGGCTTTAGAAAGATCCCCCTG 1640
DB 1201 GAGGAGCGGACAGAGGGGCGCCAGTGTCTCAGCAGAGAGGCTTTAGAAAGATCCCCCTG 1260
QY 1641 GTGCTGAGACAGACTGATGGGGCAGGGTCTGAGGATAGAGGACCGGGGAGAGGCTCTCCA 1700
DB 1261 GTGCTGAGACAGACTGATGGGGCAGGGTCTGAGGATAGAGGACCGGGGAGAGGCTCTCCA 1320
QY 1701 TGCTCATGCTCATGGCAGTACAGAGGCGCCAGGGGCGCCGGGAGGAGGCGGCGAGTCAAG 1760
DB 1321 TGCTCATGCTCATGGCAGTACAGAGGCGCCAGGGGCGCCGGGAGGAGGCGGCGAGTCAAG 1380
QY 1761 GAGTGTGTGGTCTGTGTCTAGGCTTGTGGACAGCGCGCGGCGAGCAGCTGGGCGCAAGCCGG 1820
DB 1381 GAGTGTGTGGTCTGTGTCTAGGCTTGTGGACAGCGCGCGGCGAGCAGCTGGGCGCAAGCCGG 1440
QY 1821 CTCGCTGTGTGGAGCGGCGGAGGGGTGAGCTGGAACCGCGCGCAGTCAAGAGACACTGTCAGG 1880
DB 1441 CTCGCTGTGTGGAGCGGCGGAGGGGTGAGCTGGAACCGCGCGCAGTCAAGAGACACTGTCAGG 1500
QY 1881 GAGAGGGCGAGGACAGACTGGGGGTGTGGGACAGCGCTGGGGCCCATCGGGCTATAGGGG 1940

```
Db 1501 GAGAAAGGCGAGCAGACTGGGGGTGTGGAGCAGGCGCTGGGCCCATCGGGCTATGAGGGC 1560
Qy 1941 GGGAAAGTGGGCGGGTGGGAGAGCTCGGTGGCCCTGGGCCCATCTCCAGTCCCGCCAGTCC 2000
Db 1561 GGGAAAGTGGGCGGGTGGGAGAGCTCGGTGGCCCTGGGCCCATCTCCAGTCCCGCCAGTCC 1620
Qy 2001 TGGCTCGGACAGATAGGCGGAGGTGTGCTGCTCTTTTCAGAAAGACTTCCGCGCGCACCATG 2060
Db 1621 TGGCTCGGACAGATAGGCGGAGGTGTGCTGCTCTTTTCAGAAAGACTTCCGCGCGCACCATG 1680
Qy 2061 GCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGCTGCACAGGAGCAGCAGAGC 2120
Db 1681 GCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGCTGCACAGGAGCAGCAGAGC 1740
Qy 2121 AAGCCCGAGGACAGAGGAGGAGCGCGGAGCGAGCGACCTGAGCAGCAGGCAAGGTGC 2179
Db 1741 AAGCCCGAGGACAGAGGAGGAGCGCGGAGCGAGCGACCTGAGCAGCAGGCAAGGTGC 1799

RESULT 4
ID ADJ93364 standard; DNA; 3554 BP.
AC ADJ93364;
DT 06-MAY-2004 (first entry)
XX Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.
DE
XX
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH 1695..3320
FT CDS /*tag= a
FT /*product= "Human tubulin tyrosine ligase protein"
XX
XX WO2004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX P-PSDB; ADJ93365.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
XX
```

```
PS Example 4; SEQ ID NO 12; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytosolic, respiratory
XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen.
XX antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
XX antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
XX immunosuppressive, antiseborrheic or dermatological activity acting as
XX tyrosine ligase modulators. In addition, the disclosed sequences may be
XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
XX used for diagnosing a pathological condition or a susceptibility to a
XX pathological condition in a subject, and for preventing, treating or
XX ameliorating a medical condition, such as a disorder related to aberrant
XX tubulin ligase activity, a disorder related to aberrant tubulin-
XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
XX neural disorders, brain cancer, liver cancer, or proliferative condition
XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
XX polypeptide, polynucleotide, or their modulators are also useful for
XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
XX -42 polypeptide can be used as a preventive agent for immunological
XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
XX disease or scleroderma. The antibodies may be used to purify, detect and
XX target the BGS-42 polypeptides. The present sequence is that of the
XX tubulin tyrosine ligase protein consensus gene sequence which was used in
XX the exemplification of the invention.
XX
XX Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;
XX
XX Query Match 77.8%; Score 1742.8; DB 12; Length 3554;
XX Best Local Similarity 98.3%; Pred. No. 0;
XX Matches 1784; Conservative 0; Mismatches 2; Indels 30; Gaps 1;
XX
Qy 394 GTCTAAGCGCGTGGCCAGATCACTCGGTGCCACACCCACCCAGACCCCTGTGCCCT 453
Db 1 GTCTAAGCGCGTGGCCAGATCACTCGGTGCCACACCCACCCAGACCCCTGTGCCCT 60
Qy 454 TTGCTCTGTCTCTGGTGAATCCGGGGCCCGCAGAGCTCCCTCTCAGGGGCACAGG 513
Db 61 TTGCTCTGTCTCTGGTGAATCCGGGGCCCGCAGAGCTCCCTCTCAGGGGCACAGG 120
Qy 514 CCAAGATAGGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGGGA 573
Db 121 CCAAGATAGGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGGGA 180
Qy 574 CCCAGGGGAAGAGTGCCTCCCTCCCGCAGCAGGGGCTCCCGCCCTCCCGCCCT 633
Db 181 CCCAGGGGAAGAGTGCCTCCCTCCCGCAGCAGGGGCTCCCGCCCTCCCGCCCT 240
Qy 634 CCAGCCCTGCACAAAGCAGCTTCCACACCCCTTCCCAAGGCCAGGGCTGAGGGCTG 693
Db 241 CCAGCCCTGCACAAAGCAGCTTCCACACCCCTTCCCAAGGCCAGGGCTGAGGGCTG 300
Qy 694 CCCAGGACGAGGGTGTGGGGACCTCTCTGAGGGAGGGGTCCCGGAAGAGGGGCTTCCCT 753
Db 301 CCCAGGACGAGGGTGTGGGGACCTCTCTGAGGGAGGGGTCCCGGAAGAGGGGCTTCCCT 360
Qy 754 GGACAGAGTCCCTCTCAGCAGCCAGGTGCGGCTGCCCTCAGCAGTGTGGGGCGGAG 813
Db 361 GGACAGAGTCCCTCTCAGCAGCCAGGTGCGGCTGCCCTCAGCAGTGTGGGGCGGAG 420
Qy 814 GTCAGGACAGGGTTCCTTCCGACCTAATACCCAGGGTTCAGGGCGCAGCGGCTG 873
Db 421 GTGAGGACAGGGTTCCTTCCGACCTAATACCCAGGGTTCAGGGCGCAGCGGCTG 480
Qy 874 CTAGTGGCAACATGGGCCCTTCAAAGACCCCGTGTGCAGAGCCAGTGCAGCTGCTGTG 933
Db 481 CTAGTGGCAACATGGGCCCTTCAAAGACCCCGTGTGCAGAGCCAGTGCAGCTGCTGTG 540
Qy 934 TTTTCTGGGGGCGAGCCACTTTTTTCTTAAACAGGGGTGACAAACACACATCGGGGTGCG 993
```

Db 541 TTTTCTGGGGCCAGCCACTTTTTCCTTAACGGGGTGACAAACACAGACATCGGGTGG 600
Qy 994 GGGACTTCAGATATGCTCTCTAATAGCGCGGTGCTACTCCGCTGTGGTGGGAGTGTCT 1053
Db 601 GGGACTTCAGATATGCTCTCTAATAGCGCGGTGCTACTCCGCTGTGGTGGGAGTGTCT 660
Qy 1054 GGGCATGTGGCTGTCTCGGGCCCTGACAGCGCTCTCCCTCTCTCTGGGGTCTCAGGTG 1113
Db 661 GGGCATGTGGCTGTCTCGGGCCCTGACAGCGCTCTCTCTCTCTCTGGGGTCTCAGGTG 720
Qy 1114 TGGGACACCCCTGTGGCCCTTTGGGGGCTGTGAGGAGGCCAGTCCACCGCCCTCGG 1173
Db 721 TGGGACACCCCTGTGGCCCTTTGGGGGCTGTGAGGAGGCCAGTCCACCGCCCTCGG 780
Qy 1174 TGCCCTCAGCTCAGCTCCCAATTCCTTGGGGCTGTGAGGAGGCCAGTCCACCGCCCTCAAG 1233
Db 781 TGCCCTCAGCTCAGCTCCCAATTCCTTGGGGCTGTGAGGAGGCCAGTCCACCGCCCTCAAG 840
Qy 1234 ATGGCTCTGAGCTGAGCTGCACATCCATGCTCCACCGCTGTGAGGAGGCCAGTCCACCGCCCTCAAG 1293
Db 841 ATGGCTCTGAGCTGAGCTGCACATCCATGCTCCACCGCTGTGAGGAGGCCAGTCCACCGCCCTCAAG 900
Qy 1294 CCTGGGGGACCGTATATCTGGGGAGGGGAGGATCCGCGCTAGGGGTGCTGTGTGGTG 1353
Db 901 CCTGGGGGACCGTATATCTGGGGAGGGGAGGATCCGCGCTAGGGGTGCTGTGTGGTG 960
Qy 1354 CCCATAGCCAGGCAAGTCTGAGGTGAGCGGCCCTCTCTGCTGTGTGAGGAGGCCAG 1413
Db 961 CCCATAGCCAGGCAAGTCTGAGGTGAGCGGCCCTCTCTGCTGTGTGAGGAGGCCAG 1020
Qy 1414 CTGAGCTGAGCTGCGCTCTGCTGCTTAACAGCCCACTGCTCTGAGAGTCGGGTGCTGTGTGGT 1473
Db 1021 CTGAGCTGAGCTGCGCTCTGCTGCTTAACAGCCCACTGCTCTGAGAGTCGGGTGCTGTGTGGT 1080
Qy 1474 GAACTGCGGAGCTGCGCTCTGCTGCTTAACAGCCCACTGCTCTGAGAGTCGGGTGCTGTGTGGT 1533
Db 1081 GAACTGCGGAGCTGCGCTCTGCTGCTTAACAGCCCACTGCTCTGAGAGTCGGGTGCTGTGTGGT 1140
Qy 1534 CTAGAGCTCTGCAACCGAGGTGAGCAGAGGTTCTTGGTAAAGTGAAGAGCGGAG 1593
Db 1141 CTAGAGCTCTGCAACCGAGGTGAGCAGAGGTTCTTGGTAAAGTGAAGAGCGGAG 1200
Qy 1594 AGGGCCAGTGTCTGACGAGAGGCTCTAGAAAGTCCCTCTGCTGTGAGCAGAGA 1653
Db 1201 AGGGCCAGTGTCTGACGAGAGGCTCTAGAAAGTCCCTCTGCTGTGAGCAGAGA 1260
Qy 1654 CTGATGGGGCAGGGTCTGAGGATAGAGCAGCGGAGAGGCTCCCATGCTCATGCTCAT 1713
Db 1261 CTGATGGGGCAGGGTCTGAGGATAGAGCAGCGGAGAGGCTCCCATGCTCATGCTCAT 1320
Qy 1714 GGCAGTACAGAGGCCAGGGGCCCGGGAGGGAGGGCGGCGAGTCAAGGAGTGTGTGGTTC 1773
Db 1321 GGCAGTACAGAGGCCAGGGGCCCGGGAGGGAGGGCGGCGAGTCAAGGAGTGTGTGGTTC 1380
Qy 1774 TGTGCTAGGCTTGTGACAGCGCGCGAGCAGCTGGGGCAGAGCCCGCTCCGTGTGCGGA 1833
Db 1381 TGTGCTAGGCTTGTGACAGCGCGCGAGCAGCTGGGGCAGAGCCCGCTCCGTGTGCGGA 1440
Qy 1834 GGGCCAGGGGTGACGTGAGCGGGCGCAGTACAGAGACACTGACGGGAGAGGAGGCGAGCG 1893
Db 1441 GGGCCAGGGGTGACGTGAGCGGGCGCAGTACAGAGACACTGACGGGAGAGGAGGCGAGCG 1500
Qy 1894 AGACTGGGGGTGTGGGAGCAGCCCTGGGCCCCCATCGGCTATGAGGCGGGAGTGGGGCG 1953
Db 1501 AGACTGGGGGTGTGGGAGCAGCCCTGGGCCCCCATCGGCTATGAGGCGGGAGTGGGGCG 1560
Qy 1954 GGTGGGAGGCTCCGTGGCCCTGCCCCCATCTCCAGTCCCGAGTCTGGCTCGGACAGA 2013
Db 1561 GGTGGGAGGCTCCGTGGCCCTGCCCCCATCTCCAGTCCCGAGTCTGGCTCGGACAGA 1620
Qy 2014 TAGGCGAGGCTGTCTGTCTTCA-----GAAG 2043

Db 1621 TAGGCGAGGCTGTCTGTCTCTTTCAGAGTAGCAGCGTGGGAAGAAGGAGTCTCTGGAAG 1680
Qy 2044 ACTTCCGGCGCACCATGTCATCCAGCATCTCTCAAGTGGTGGTCAAGCCACAGAGCTGCA 2103
Db 1681 ACTTCCGGCGCACCATGTCATCCAGCATCTCTCAAGTGGTGGTCAAGCCACAGAGCTGCA 1740
Qy 2104 GCAGGAGCAGCAGCAAGCAAGCCAGGAGCCAGAGGAGGAGCGCCGGAGCAGCGACCTGA 2163
Db 1741 GCAGGAGCAGCAGCAAGCAAGCCAGGAGCCAGAGGAGGAGCGCCGGAGCAGCGACCTGA 1800
Qy 2164 GCAGGAGCAGCAAGGTGC 2179
Db 1801 GCAGGAGCAGCAAGGTGC 1816
RESULT 5
ADJ93357
ID ADJ93357 standard; cDNA; 1838 BP.
XX
AC ADJ93357;
XX
DT 06-MAY-2004 (first entry)
XX
Human BGS-42 cDNA sequence SeqID1.
XX
testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteoprotective; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.
XX
Homo sapiens.
XX
Key Location/Qualifiers
CDS 153..1778
FT /*tag= a
FT /product= "Human BGS-42 protein"
XX
WO2004005487-A2.
XX
15-JAN-2004.
XX
09-JUL-2003; 2003WO-US021605.
XX
09-JUL-2002; 2002US-0394725P.
XX
(BRIM) BRISTOL-MYERS SQUIBB CO.
XX
Feder JN, Wu S, Nelson TC;
XX
WPI; 2004-099381/10.
XX
P-PSDB; ADJ93358.
XX
New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
useful for preventing, treating or ameliorating a medical condition, e.g.
aberrant cellular proliferation, reproductive disorders or testicular
disorders.
XX
Claim 1; SEQ ID NO 1; 343pp; English.
XX
This invention relates to a novel testis-specific tubulin tyrosine-ligase
-like polypeptide, designated the BGS-42 polypeptide. The invention may
be useful for the development of compounds with a cytostatic, respiratory
-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,

CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.

XX SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Query Match 12.2%; Score 272.4; DB 12; Length 1838;
Best Local Similarity 99.6%; Pred. No. 2.7e-49;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCATCCAGTCCCGGCGCACATCGGATCGGATGGGCGAGCT 2025
Db 1 TGGGAGCAGGCTGGGCCCCATCCAGTCCCGGCGCGAAAGTGCGGCGGTTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCATCCAGTCCCGGCGCACATCGGATCGGATGGGCGAGCT 2025
Db 61 TCCGTGGCCCTGGGCCCCATCCAGTCCCGGCGCGAAAGTGCGGCGGTTGGGGAGCT 120

Qy 2026 GTGCTGTCTTTTTCAGAGACTTCCTGGGCGCACATCGGATCGGATGGGCGAGCT 2085
Db 121 GTGCTGTCTTTTTCAGAGACTTCCTGGGCGCACATCGGATCGGATGGGCGAGCT 180

Qy 2086 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCAGGAGCAGGAGGAGG 2145
Db 181 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCAGGAGCAGGAGGAGG 240

Qy 2146 CCGGGAGCAGGACTGCAGCAGGAGCAGGAGTGC 2179
Db 241 CCGGGAGCAGGACTGCAGCAGGAGCAGGAGTGC 274

RESULT 6

ADJ93361

ID ADJ93361 standard; DNA; 1939 BP.

XX AC ADJ93361;

XX AC ADJ93361;

DT 06-MAY-2004 (first entry)

XX Human BGS-42 protein-related DNA clone A SeqID9.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;

KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.

XX Homo sapiens.

OS Homo sapiens.

PN WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Example 4; SEQ ID NO 9; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytosolic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen.
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.

SQ Sequence 1939 BP; 421 A; 592 C; 589 G; 337 T; 0 U; 0 Other;

Query Match 12.2%; Score 272.4; DB 12; Length 1939;

Best Local Similarity 99.6%; Pred. No. 2.7e-49;

Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCATCGGATCGGATGGGCGAGCTGGGCGAGCT 1965
Db 1 TGGGAGCAGGCTGGGCCCCATCGGATCGGATGGGCGAGCTGGGCGAGCT 60

Qy 1966 TCCGTGGCCCTGGGCCCCATCCAGTCCCGGCGCACATCGGATCGGATGGGCGAGCT 2025
Db 61 TCCGTGGCCCTGGGCCCCATCCAGTCCCGGCGCACATCGGATCGGATGGGCGAGCT 120

Qy 2026 GTGCTGTCTTTTTCAGAGACTTCCTGGGCGCACATCGGATCGGATGGGCGAGCT 2085
Db 121 GTGCTGTCTTTTTCAGAGACTTCCTGGGCGCACATCGGATCGGATGGGCGAGCT 180

Qy 2086 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCAGGAGCAGGAGGAGG 2145
Db 181 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCAGGAGCAGGAGGAGG 240

QY 2146 CCGGAGCAGCAGCCTGAGCAGCAGCAGCAGGTGC 2179
|||||
Db 241 CCGGAGCAGCAGCCTGAGCAGCAGCAGGTGC 274
|||||

RESULT 7
ADJ93362
ID ADJ93362 standard; DNA; 1859 BP.
XX
AC ADJ93362;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone B SeqID10.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory; antiinflammatory; gastrointestinal; antiinflammatory; anabolic; hypotensive;
KW neuroprotective; endocrine; antiinflammatory; antiinflammatory; anabolic; hypotensive;
KW osteopathic; neurotropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine kinase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
PN WO2004005487-A2.
XX
PD 15-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
DR WPI; 2004-099381/10.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
PS
PS Example 4; SEQ ID NO 10; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytosolic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine kinase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for

CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;
Query Match 6.2%; Score 138.4; DB 12; Length 1859;
Best Local Similarity 99.3%; Pred. No. 3.8e-20;
Matches 139; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2040 GAAGACTTCGGCGCACCATGGCATTCAGATCTCTCAAGTGGTGGTGCAGCAGCAGC 2099
Db 29 GAAGACTTCGGCGCACCATGGCATTCAGATCTCTCAAGTGGTGGTGCAGCAGCAGC 88
QY 2100 TGCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2159
Db 89 TGCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 148
QY 2160 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2179
Db 149 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 168

RESULT 8
ADQ64426
ID ADQ64426 standard; cDNA; 5282 BP.
XX
AC ADQ64426;
XX
DT 07-OCT-2004 (first entry)
XX
DE Novel human cDNA sequence #1587.
XX
KW ss; gene; osteopathic; neuroprotective; neurotropic; antiparkinsonian;
KW cytosolic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW cancer.
XX
OS Homo sapiens.
XX
PN EP1440981-A2.
XX
PD 28-JUL-2004.
XX
PF 21-JAN-2004; 2004EP-00001196.
XX
PR 21-JAN-2003; 2003JP-00102206.
PR 09-MAY-2003; 2003JP-00131392.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Nagai K, Irie R;
DR WPI; 2004-535376/52.
DR P-PSDB; ADQ66614.
XX
PT Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS Claim 1; SEQ ID NO 1587; 2449pp; English.
XX
XX The invention relates to 2495 novel polynucleotides (I) and their encoded
CC polypeptides, sequences hybridizing to these nucleotides, sequences
CC encoding partial polypeptides and sequences having 70% or 90% identity to
CC the nucleotide and protein sequences. The nucleotides and polypeptides
CC are useful as diagnostic markers or therapeutic target for the diseases
CC or morbid states. They are also useful for treating osteoporosis,
CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1042 BP; 227 A; 332 C; 288 G; 195 T; 0 U; 0 Other;

Query Match 3.0%; Score 67.8; DB 5; Length 1042;
Best Local Similarity 62.9%; Pred. No. 7.5e-05;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 96 TGAATGCTGGGCATCAATAAAGACTTTTCTTCCAGCTAGCTGGTAAATAATGAGA 155
DB 707 TGAATTTTGATGACCTAGATGGAACACATGCTCTGATGTCGCGCATGTCACGAATGAGA 766
QY 156 TCCCTACTCTCTGGACCATCAAGAGGACATCATTTGACCTATCAGCTGACCTACG 215
DB 767 TCCCTACTCTATCTGGACCATCGCGGGATGCTGCTGCTGCTCTCTCTCAAGG 826
QY 216 ACCAGATGCTGAACCACTACGCAAGACAGCGCTCTCTTCCACCAAG 262
DB 827 ATCAGATGATAAACCACTACGCGCGGTGGCTCTTTTACCACAAAG 873

RESULT 11
AAS69383
ID AAS69383 standard; cDNA; 2250 BP.

XX AAS69383;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #5187.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US008631.
XX
XX PR 31-MAR-2000; 2000US-00540217.
XX
XX PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-FSDB; ABG05196.

New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

PS Claim 1; SEQ ID NO 5187; 103pp; English.

XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 2250 BP; 699 A; 558 C; 620 G; 373 T; 0 U; 0 Other;

Query Match 3.0%; Score 67.8; DB 5; Length 2250;
Best Local Similarity 62.9%; Pred. No. 8.9e-05;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 96 TGAATGCTGGGCATCAATAAAGACTTTTCTTCCAGCTAGCTGGTAAATAATGAGA 155
DB 1427 TGAATTTTGATGACCTAGATGGAACACATGCTCTGATGTCGCGCATGTCACGAATGAGA 1486
QY 156 TCCCTACTCTCTGGACCATCAAGAGGACATCATTTGACCTATCAGCTGACCTACG 215
DB 1487 TCCCTACTCTATCTGGACCATCGCGGGATGCTGCTGCTGCTCTCTCTCAAGG 1546
QY 216 ACCAGATGCTGAACCACTACGCAAGACAGCGCTCTCTTCCACCAAG 262
DB 1547 ATCAGATGATAAACCACTACGCGCGGTGGCTCTTTTACCACAAAG 1593

RESULT 12

ADC86736
ID ADC86736 standard; DNA; 5452 BP.
XX
XX ADC86736;
XX
XX 01-JAN-2004 (first entry)
XX
XX Human GPCR gene SEQ ID NO:1189.

XX ds; gene; human; GPCR;
XX guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
XX Homo sapiens.

XX EP1270724-A2.

XX 02-JAN-2003.

XX 18-JUN-2002; 2002EP-00013517.

XX 18-JUN-2001; 2001JP-00246789.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.

XX Suwa M, Asai K, Akiyama Y, Aburatani H;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:08:28 ; Search time 8612.22 Seconds
(without alignments)
9904.766 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgcctgcagcagaccggc.....gcctccacagggccctccc 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gse1:*

9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	137.4	6.1	1259	9	AY415398	Homo sapi
2	135.8	6.1	1257	9	AY415399	Pan trogl
3	129	5.8	129	7	H55115	CHR220054 C
4	111.8	5.0	213	7	H55109	CHR220048 C
5	97	4.3	2969	3	AK030151	Mus muscu
6	93.8	4.2	487	8	A2248094	RPCT-23-4
7	93.8	4.2	2720	3	AK029745	Mus muscu
8	83.2	3.7	1218	9	CF081011	CH216-160
9	76.4	3.4	467	2	BF903338	PMI-MT019
10	74	3.3	623	6	BF745669	BY745669
11	74	3.3	2520	3	AK080321	Mus muscu
12	73.4	3.3	1174	6	CD504184	CD466-E10
13	73.2	3.3	1585	8	CC294035	CH261-82A
14	71.6	3.2	3328	3	BC028169	Homo sapi
15	71.4	3.2	1040	9	CNS030AX	AL253122 Tetraodon
16	70.8	3.2	560	9	CG646452	OST192793
17	70.8	3.2	741	7	CNS26715	UI-M-HB0-
18	70.8	3.2	1451	3	AK029462	Mus muscu
19	70.6	3.2	582	5	BP292995	BP292995
20	70.6	3.2	1315	9	AG279688	Mus muscu
21	70.6	3.2	1909	9	CL078604	CH216-151
22	70	3.1	252	9	CG546562	OST146147
23	70	3.1	925	9	CNS0091P	Drosophila
24	69.8	3.1	581	5	BP357921	BP357921

25	69.6	3.1	1069	9	CL118121	ISB1-70B1
26	69.4	3.1	1137	9	AG078502	Pan trogl
27	69.2	3.1	251	9	CG592880	OST249456
28	69.2	3.1	251	9	CG599359	OST266453
29	69.2	3.1	378	9	CG624680	OST329018
30	69.2	3.1	452	9	CG532512	OST116474
31	69	3.1	910	8	BH162948	ENTRKA9TR
32	68.8	3.1	525	7	CF918454	B0995D07-
33	68.6	3.1	519	2	AW656035	108059 MA
34	68.6	3.1	573	2	AW656034	108058 MA
35	68.6	3.1	825	7	CK770333	958489 MA
36	68.4	3.1	821	9	CNS049PN	Tetraodon
37	68.4	3.1	891	9	CL477366	SAIL_271
38	68.4	3.1	1038	9	CL478560	SAIL_293-
39	68.2	3.0	1340	9	AG430438	Mus muscu
40	68.2	3.0	1405	9	CL509298	SAIL_811
41	68	3.0	1690	9	AG435238	Mus muscu
42	68	3.0	1767	8	CC294031	CH261-82A
43	67.8	3.0	1153	8	CC252982	CH261-187
44	67.8	3.0	1269	5	BM914457	AGENCOURT
45	67.8	3.0	1523	9	AG448267	Mus muscu

ALIGNMENTS

RESULT 1
AY415398
LOCUS
DEFINITION Homo sapiens HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY415398
VERSION AY415398.1 GI:39771357
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1259)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1259)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source 1..1259
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1259
/locus_tag="HCM5544"

Query Match 6.1%; Score 137.4; DB 9; Length 1259;
Best Local Similarity 99.3%; Pred.No.2.6e-19;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2041 AAGACTTCCGCGCAGCATGCCATCCAGCATCCTCAAGTGGTGGTCCAGCCAGCAGCT 2100
Db 1 AAGACTTCCGCGCAGCATGCCATCCAGCATCCTCAAGTGGTGGTCCAGCCAGCAGCT 60

```

QY      2101 GCAGCAGGAGCAGCAGAAGCAAGCCAGGAGCCAGGAGCCAGAGGGAGGAGCGCGGAGCGAGCGAC 2160
Db      61 GCAGCAGGAGCAGCAGAAGCAAGCCAGGAGCCAGGAGCCAGAGGGAGGAGCGCGGAGCGAGCGAC 120

QY      2161 TGAGCAGCAGCGAAGGTGC 2179
Db      121 TGAGCAGCAGCGAAGGTGC 139

RESULT 2
AY415399
LOCUS   1257 bp DNA linear GSS 12-DEC-2003
DEFINITION Pan troglodytes HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION AY415399
VERSION 1
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 1257)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1257)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source
1. .1257
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1257
/locus_tag="HCM5544"
gene
ORIGIN
Query Match 6.1%; Score 135.8; DB 9; Length 1257;
Best Local Similarity 98.6%; Pred. No. 5.9e-19;
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2041 AAGACTTCCGGCGCACCATGCGATCCAGCATCTCTCAAGTGCGTGCGTCAAGCCACAGAGCT 2100
Db      1 AAGACTTCCGGCGCACCATGCGATCCAGCATCTCTCAAGTGCGTGCGTCAAGCCACAGAGCT 60

QY      2101 GCAGCAGCAGCAGAGCAAGCCAGGAGCCAGAGGGAGGAGCGCGGAGCGAGCGAC 2160
Db      61 GCAGCAGGAGCGGAGAGCAAGCCAGGAGCCAGAGGGAGGAGCGCGGAGCGAGCGAC 120

QY      2161 TGAGCAGCAGCGAAGGTGC 2179
Db      121 TGAGCAGCAGCGAAGGTGC 139

RESULT 3
H55115
LOCUS   129 bp mRNA linear EST 07-DEC-1995
DEFINITION CHR220054 Chromosome 22 exon Homo sapiens cDNA clone C22_74 5', mRNA sequence.
ACCESSION H55115

```

```

VERSION H55115.1 GI:1107981
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 129)
AUTHORS Trofatter,J.A., Long,K.R., Murrell,J.R., Stotler,C.J., Gusella,J.F. and Buckler,A.J.
TITLE An expression-independent catalog of genes from human chromosome 22
JOURNAL Genome Res. 5 (3), 214-224 (1995)
MEDLINE 96159527
PUBMED 8593609
COMMENT Contact: Buckler AJ
Molecular Neurogenetics Unit
Massachusetts General Hospital
Building 149, 13th St., Charlestown MA 02129
Tel: 6177249616
Fax: 6177265736
Email: buckler@helix.mgh.harvard.edu
Seq primer: T3.
FEATURES
source
1. .129
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="C22_74"
/lab_host="E. coli DH5a"
/clone_lib="Chromosome 22 exon"
/notes="Vector: pBluescriptIIKS+; Site 1: Sal I; Site 2: Bam HI (destroyed); Exons were isolated from human chromosome 22 specific cosmids using a modification of the method of exon amplification (Proc. Natl. Acad. Sci. USA 88:4005-4009, 1991). Amplified exons were digested with Sal I and Bgl II and subsequently cloned into pBluescriptIIKS+ at the Sal I and Bam HI sites."
ORIGIN
Query Match 5.8%; Score 129; DB 7; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.4e-17;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      134 TCTAGGTTGGTAAATAATGAGATGCCGTACCTCTCTGGACCATCAAGAGGACATCATT 193
Db      1 TCTAGGTTGGTAAATAATGAGATGCCGTACCTCTCTGGACCATCAAGAGGACATCATT 60

QY      194 GACTATCACAGCTGACCTACGACCATGCTGAACCACTACGCAAAAGACAGCCTCCTTC 253
Db      61 GACTATCACAGCTGACCTACGACCATGCTGAACCACTACGCAAAAGACAGCCTCCTTC 120

QY      254 ACCACCAAG 262
Db      121 ACCACCAAG 129

RESULT 4
H55109
LOCUS   213 bp mRNA linear EST 07-DEC-1995
DEFINITION CHR220048 Chromosome 22 exon Homo sapiens cDNA clone C22_67 5', mRNA sequence.
ACCESSION H55109
VERSION 1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 213)
AUTHORS Trofatter,J.A., Long,K.R., Murrell,J.R., Stotler,C.J., Gusella,J.F. and Buckler,A.J.
TITLE An expression-independent catalog of genes from human chromosome 22
JOURNAL Genome Res. 5 (3), 214-224 (1995)
MEDLINE 96159527

```

8593609
 CONTACT: Buckler AJ
 Molecular Neurogenetics Unit
 Massachusetts General Hospital
 Building 149, 13th St., Charlestown MA 02129
 Tel: 6177249616
 Fax: 6177265736
 Email: buckler@helix.mgh.harvard.edu
 Seq primer: 73.

FEATURES
 source
 1..213
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="C22-67"
 /lab_host="E. coli DH5a"
 /clone_lib="Chromosome 22 exon"
 /note="Vector: pBluescriptIIKS+; Site 1: Sal I; Site 2: Bam HI (destroyed); Exons were isolated from human chromosome 22 specific cosmids using a modification of the method of exon amplification (Proc. Natl. Acad. Sci. USA 88:4005-4009, 1991). Amplified exons were digested with Sal I and Bgl II and subsequently cloned into pBluescriptIIKS+ at the Sal I and Bam HI sites."

ORIGIN
 Query Match 5.0%; Score 111.8; DB 7; Length 213;
 Best Local Similarity 98.3%; Pred. No. 8.9e-14;
 Matches 113; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1460 ATCGGGCTGTGCTGAACATGCGAGCTGCGCTGCTAGCTCCGCGCAACCCGACTCC 1519
 |||||||
 Db 1 ATCGGGCTGTGCTGAACATGCGAGCTGCGCTGCTAGCTCCGCGCAACCCGACTCC 60
 |||||||

QY 1520 TTCTTCCACGCTGTACAGCTCTGACCGAGAGTGAGCAGAGTCTCTCTGG 1574
 |||||||
 Db 61 TTCTTCCACGCTGTACAGCTCTGACCGAGAGTGAGCAGAGTCTCTCTGG 115
 |||||||

RESULT 5
 AK030151
 LOCUS
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933401B17 product:hypothetical tubulin-tyrosine ligase containing protein, full insert sequence.
 AK030151
 ACCESSION
 VERSION AK030151.1 GI:26326134
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P., and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

REFERENCE
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374

REFERENCE
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6 (bases 1 to 2969)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 Location/Qualifiers
 1..2969
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:4933401B17"
 /db_xref="taxon:10090"
 /clone="4933401B17"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 490..2835
 /note="unnamed protein product; hypothetical tubulin-tyrosine ligase containing protein (InterPro|IPR004344, evidence: InterPro) putative"
 /codon_start=1
 /protein_id="BAC26811.1"
 /db_xref="GI:26326135"
 /translation="MENGKSLSTLSDGDHKEENKLGKIPDLSSPKLDYKIA RQUTEAKERKIFSIYGHYPVIRATLRKGNWVEKFNFPKALQNLGSEKSAETKE KIGLAEFRFDIHDVMSRLVKNPEIPYLLWTIKRDVDYHSLTCDMLNHVYKTSFTQ NQIENLMSRLPWYQANPTFFRCQSEKQFLDQFTVAATSLKVVYLVHQ NDCSKVKGSKSEAKNSDPSPKDPENPDLKLPSPGVVDVTKQVQAYLGQLEHE DIVDSEASTALSEEWNDLTQOYVLLVHGNASITDSKSYFAOQALLSKISSVNPOT EIDGIRNIWIKPAKSRGRDVCMDRVENILSLVAADSTQTKDNKVVQYK1ETPML

FEATURES
 source
 1..2969
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:4933401B17"
 /db_xref="taxon:10090"
 /clone="4933401B17"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 490..2835
 /note="unnamed protein product; hypothetical tubulin-tyrosine ligase containing protein (InterPro|IPR004344, evidence: InterPro) putative"
 /codon_start=1
 /protein_id="BAC26811.1"
 /db_xref="GI:26326135"
 /translation="MENGKSLSTLSDGDHKEENKLGKIPDLSSPKLDYKIA RQUTEAKERKIFSIYGHYPVIRATLRKGNWVEKFNFPKALQNLGSEKSAETKE KIGLAEFRFDIHDVMSRLVKNPEIPYLLWTIKRDVDYHSLTCDMLNHVYKTSFTQ NQIENLMSRLPWYQANPTFFRCQSEKQFLDQFTVAATSLKVVYLVHQ NDCSKVKGSKSEAKNSDPSPKDPENPDLKLPSPGVVDVTKQVQAYLGQLEHE DIVDSEASTALSEEWNDLTQOYVLLVHGNASITDSKSYFAOQALLSKISSVNPOT EIDGIRNIWIKPAKSRGRDVCMDRVENILSLVAADSTQTKDNKVVQYK1ETPML

CDS

IYTKFDIROWFLVTDNPLTITWIFYKESYLRFSTQSFSLDKLSAHLNNSIORLKL
NDKERSPLLCNNMTSTRFQYLQKRGRTGWSIIYPSMKRAVTNAMEVAQDHVEA
RKNSFELYGADFILGRDFKWLJLEINSSPTMHSTPTVAQLCAQVEDITIKVVVDRL
DRNCIDGNPAVELPFPNGSDLCVEGIVSKAKKOMPPIASVGLSRLSDAP
PKYRSALMETVIRPRTTIVRODMKREEAJVLSTWSPMDVAEVRGRAPKIYAEV
NDQHDVNDKSHKGYTRVQSSKVPFVTLTSAQHPALFAQTMKFTQMTSSPPTASGNH
RDSSPFCPIVFEELWHLPHNSQRRPSSCILQSAQGWIRGIP"

ORIGIN

Query Match 4.3%; Score 97; DB 3; Length 2969;
Best Local Similarity 81.8%; Pred. No. 2.2e-10; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 25;

QY 133 GTCTAGTTTGGTAAAAATGAGATGCGGTACCTCTCTGGACCATCAAGAGGACATCAT 192
|||||
Db 843 GTCCAGTTTGGTAAAGAAGAGATCCATCTCTCTGGACTATCAAGAGGATGTTGT 902
|||||

QY 193 TGACTATCAGCCTGACCTAGCAGCAGATGCTGAACCATCTAGCAAGAAGACAGCTCTT 252
|||||
Db 903 GGACTACACAGCCTGACCTGTGACAGATGCTAAATCTACCGGGAAGACAGCATCAT 962
|||||

QY 253 CACCACCAAGGTGAGCC 269
|||||
Db 963 CACAACCAAGATTGGAC 979
|||||

RESULT 6
AZ248094/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AZ248094 487 bp DNA linear GSS 15-JUN-2000
RPCI-23-41G15-TV RPCI-23 Mus musculus genomic clone RPCI-23-41G15,
Genomic survey sequence.
A248094
A2248094.1 GI:8561297
GSS.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 487)
Zhao S., Nierman W., Feldblyum T., Malek J., Shatsman S.,
Akinret B., Levins M., Megann S., Tsegaye G., Geer K., Krol M., de
Jong P. and Frazer C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.bufo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.bufo.edu/orderingframe.html>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac ends/mouse/bac_end_intro.html
Plate: 41 row: G column: 15
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. .487
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-41G15"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/notes="Organ: Kidney/Brain; Vector: pBACe3 6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested"

FEATURES
source

1. .487
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-41G15"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/notes="Organ: Kidney/Brain; Vector: pBACe3 6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested"

with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "

ORIGIN

Query Match 4.2%; Score 93.8; DB 8; Length 487;
Best Local Similarity 82.9%; Pred. No. 8.8e-10;
Matches 107; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 137 AGTTTGGTAAAAATGAGATGCGGTACCTCTCTGGACCATCAAGAGGACATCATTCAC 196
|||||
Db 483 AGTTTGGCAAGAACAGATCCCATCTCTCTGGACTATCAAGAGGATGTTGGAC 424
|||||

QY 197 TATCACAGCTGACCTACGACAGATGCTGAACCATCTAGCAAGAAGACAGCTCTCTTCA 256
|||||
Db 423 ACCACAGCTGACCTGTGACAGATGCTAAATCTACCTACGGGAAGACAGCATTCACA 364
|||||

QY 257 ACCAAGGTG 265
|||||
Db 363 ACCAAGGTG 355
|||||

RESULT 7
AK029745
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AK029745 2720 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4930524K07 product:hypothetical Tubulin-tyrosine
ligase containing protein, full insert sequence.
AK029745.1 GI:26325681
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakeguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5

The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

JOURNAL
REFERENCE
AUTHORS
Nature 420, 563-573 (2002)
6 (bases 1 to 2720)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohmoto, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
source
1..2720
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:4930524K07"
/db_xref="taxon:10090"
/clone="4930524K07"
/sex="male"
/tissue type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"
934..2490
/note="unnamed protein product; hypothetical Tubulin-tyrosine ligase containing protein (InterPro|IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26595.1"
/db_xref="GI:26325682"
/translation="MLNHYGKTASFTTKIGLCLNMRSLPWYVQANPNTFFPRCYGLCT ESEKQEFLLDFFRTVAASILKVVVLHQNVCYKVKGKKEAKNSDPSKDPENPDL KLPSLGGVDTACKVCOAYLQLEHEDI DVSEASTEALSEEWNLDLTQOYLLVHGN ASLTDKSYFAQCQALLSKISVNPQTEIDGIRNIWLIKPAKSRGRDVCMDREVENI LSTVADSQTTKDNKVVQKVIETPMLIYDTKFDIRQWFLVTDNPLTIFPKESYLRL FSVRAUQKLDLSAIHLNNSIQRLKNDKERSPLLPCHNMTSTRFOYLQKRGGRG TWGSIITPSMRVATNMRVADQHVAKNSFELYGADFLGRDFKPLILINSPTM HPSPVTAQLCAQVQEDITIKVVDRKLDKNDICIGNFELMRQPAVELPPFNGSOLCWE GISVYKAKKQMPPIASVGLSESLDAPPKVRSARALMETVIRPRTTVRQDKREEAK GFRVISKERA"
ORIGIN
Query Match 4.2%; Score 93.8; DB 3; Length 2720;
Best Local Similarity 80.3%; Pred. No. 1.1e-09;
Matches 110; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 133 GTCTAGGTTGTTAAAAATGAGATCGCTACTCTCTGGACCATCAAGAGGACATCAT 192
Db |||||
846 GTGATGGTTGTTAAAGACGAGATCCCATACCTCTCTGGACTATCAAGAGGAGTGTGT 905
Qy |||||
193 TGACTATCACAGCCTGACCTAGCACCATCTGAACCACTACGCAAGACAGCCTCCTT 252
Db |||||
906 GGACTACCACAGCCTGACCTGTGTCAGATGCTAAATCACTACGGGAAGACAGCATCAT 965

Qy 253 CACCACCAAGGTGAGCC 269
Db 966 CACAACCAAGATTGGAC 982

RESULT 8
CL081011/c 1218 bp DNA linear GSS 31-DEC-2003
LOCUS CH216-160116_RMA.1 CH216 Xenopus tropicalis genomic clone
DEFINITION CH216-160116, genomic survey sequence.
ACCESSION CL081011
VERSION CL081011.1 GI:405336924
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1218)
AUTHORS Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM4 ctcaagggtcgtcgagc
Class: BAC ends
High quality sequence start: 79
High quality sequence stop: 190.
Location/Qualifiers
1..1218
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-160116"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pPARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN
Query Match 3.7%; Score 83.2; DB 9; Length 1218;
Best Local Similarity 41.7%; Pred. No. 2.1e-07;
Matches 286; Conservative 0; Mismatches 400; Indels 0; Gaps 0;
Qy 1550 GAGAGTGAGCAGCAGGAGTTCTCTGGGTAAGTGAGGAGACGCGCAGAGGCGCCAGTGCTGT 1609
Db 986 GGGGGGGCGGG 927
Qy 1610 CAGCAGAGAGGCTTCTTAAAGAATCCCTCTGGTCTCTAGACAGACTGATGGGGCAGGTC 1669
Db 926 GGGGGGGCGGG 867
Qy 1670 TGAGATAGAGGACCGGGGAGAGCCCTCCCATGTGTCATGTGTCATGGCAGTACAGAGGCCA 1729
Db 866 CCGGG 807
Qy 1730 GGGGGCCCGGGAGGAGGCGGGCGGCGAGTCAAGAGAGTGTGTGTCTGTGTCTAGGCTTGTGG 1789
Db 806 GCGCGCGCGGG 747
Qy 1790 ACAGCGCGGCGAGCAGCTGGGGGCAAGCCCGGCTCCGTGTCTCGAGAGCCCGAGGGGTGACG 1849
Db 746 GCGGGGGCGGG 687
Qy 1850 TGGACGGCCCGCAGTCACAGACACTCGACGAGAGAGGCGAGGAGGAGGAGGAGGAGGAGG 1909
Db |||||

The Institute of Physiscal and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayaahida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Konno, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES

Location/Qualifiers
1..623
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630053H17"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DHL0B"
/clone_lib="RIKEN full-length enriched, 3 days neonate
thymus"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5' GAGAGAGATTCTCGAGTTAATTAATTAATTCCTCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

ORIGIN

Query Match 3.3%; Score 74; DB 6; Length 623;
Best Local Similarity 69.2%; Pred. No. 2e-05;
Matches 101; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
124 TTCTTGGCAGTCTAGGTGGTAAATATGATCCGCTACCTCTCTGGACATCAAGAG 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 TTCTTCCAGTCCCGATGGTTCCGAATGAGACCCCTACCTCTCTGACCCTCGCG 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 184 GGACATCATTTGACTATCACAGCCTGCTAGCAGCATCTGACCACTAGCAAGAC 243
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 GGATGCTGATGTGCTGCTCTTTCCAGGATCAGATGATAAACCATTATGCCGTGC 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 244 AGCCTCTCTTACCACCAAGGTGAGCC 269
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 AGGCTCTCTTACCACCAAGGTGAGCC 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11

AK080321 2520 bp mRNA linear HTC 03-APR-2004
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:A630053H17 product:HOTTL PROTEIN homolog [Homo
sapiens], full insert sequence.
AK080321
AK080321.1 GI:26348476
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayaishizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

FEATURES
source
1. .2520
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630053H17"
/db_xref="taxon:10090"
/clone="A630053H17"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
84. .2198
[note="unnamed protein product; HOTTL PROTEIN homolog (Homo sapiens) [SPTR|Q5UI99, evidence: PASTY, 85.8%ID, 93.4%length, match=821]
putative"]
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"
/translations="MDTSSPFYAGGCEASPPSLPVSISPSLQSRMVRNETPVLWT
TRDVLDCRFLSKQDMNHVARAGSFYTKVGLCLNLRNLPWFDEADSPFPCRYRLG
AEDKKAIFDFMLTAAKNVLRKLVKLEKSSQISIQAREEAPEDTPKQKBLVT
VSSDFVEALSAQOHLSSIAHKDIDKPNPLSLPDDWSQFLQRYQIVHGBELR
YLSVFQRCDEILQONVVPQIDMDGRNIWIKFGAKSRGRGIMCMRLDMLKLV
DCNPLMKDGKWTQVKYIERPLIFGTGFDLRQWFLVDMNPLTVMFYRDSYIRFSTQ
PFSLKNLDNSVHLCCNSIORHLEASCHRHPLPDDNWSORFOAHLEQVDAFKWSS
VIVPGKAAVIALQTSQDNVQCRKASFELYGADFVGEDEFQWLEINASPMPAST
AVTARLCAGVQADTLRVVIDRRLDRSCDTGAFELITYQPAVEVPQIVGIRLLVEGSTI
KKPVPVGHRRGTGVRSLPHLLTQGGESKDSGSPHRSASRNKRAESLHTEKPEP
AAVASVGKGGKAPFPFSLHSAWLFPSPRVHRPQGRVLRQLQDLVGSKALSTGKA
LMTLPTAKVLMSPPHPDLKLAPSLMKPGKVGPELCTTWRVVLSSGGIGBEGRQAA
PRSSAPGKGLSSTPCKSTET"

CDS
[Homo sapiens] [SPTR|Q5UI99, evidence: PASTY, 85.8%ID, 93.4%length, match=821]
putative"
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"
/translations="MDTSSPFYAGGCEASPPSLPVSISPSLQSRMVRNETPVLWT
TRDVLDCRFLSKQDMNHVARAGSFYTKVGLCLNLRNLPWFDEADSPFPCRYRLG
AEDKKAIFDFMLTAAKNVLRKLVKLEKSSQISIQAREEAPEDTPKQKBLVT
VSSDFVEALSAQOHLSSIAHKDIDKPNPLSLPDDWSQFLQRYQIVHGBELR
YLSVFQRCDEILQONVVPQIDMDGRNIWIKFGAKSRGRGIMCMRLDMLKLV
DCNPLMKDGKWTQVKYIERPLIFGTGFDLRQWFLVDMNPLTVMFYRDSYIRFSTQ
PFSLKNLDNSVHLCCNSIORHLEASCHRHPLPDDNWSORFOAHLEQVDAFKWSS
VIVPGKAAVIALQTSQDNVQCRKASFELYGADFVGEDEFQWLEINASPMPAST
AVTARLCAGVQADTLRVVIDRRLDRSCDTGAFELITYQPAVEVPQIVGIRLLVEGSTI
KKPVPVGHRRGTGVRSLPHLLTQGGESKDSGSPHRSASRNKRAESLHTEKPEP
AAVASVGKGGKAPFPFSLHSAWLFPSPRVHRPQGRVLRQLQDLVGSKALSTGKA
LMTLPTAKVLMSPPHPDLKLAPSLMKPGKVGPELCTTWRVVLSSGGIGBEGRQAA
PRSSAPGKGLSSTPCKSTET"

ORIGIN
Query Match 3.3%; Score 74; DB 3; Length 2520;
Best Local Similarity 69.2%; Pred. No. 2.4e-05;
Matches 101; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 124 TTCTTCCAGCTAGGTTGGTAAATAATGAGATCGCTACTCTCTGGACCATCAAGAG 183
DB 164 TTCTCTCAGTCCCGCATGGTTCGAAATGAGACCCCTACTCTATCTGGACCATCGGCG 223
QY 184 GGACATCATTTGACTATCACAGCTCGACCTACGACCATGCTGAACCACTACGCAAGAC 243
DB 224 GGATGCTGATTTGCTTCTCTTCCAAAGATCAGATGATATAACCATTTATGCCCGTGC 283
QY 244 AGCCTCTTCCACCAAGGTGAGCC 269
DB 284 AGGCTCTTCCACCAAGGTGGGCC 309

RESULT 12
CD504184

LOCUS CD504184 1174 bp mRNA linear EST 12-JUN-2003
DEFINITION CDA66-E10.x1d-t SRGC-CDA Gasterosteus aculeatus cDNA clone
CDA66-E10 5', mRNA sequence.
ACCESSION CD504184
VERSION CD504184.1 GI:31434752
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1174)
Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J. and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@crgm.stanford.edu
Plate: 66
High quality sequence stop: 289.

FEATURES
source
1. .1174
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA66-E10"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

ORIGIN
Query Match 3.3%; Score 73.4; DB 6; Length 1174;
Best Local Similarity 28.7%; Pred. No. 3e-05;
Matches 206; Conservative 0; Mismatches 513; Indels 0; Gaps 0;
QY 1483 GAGCCCTGCCCTGGTACGTCCCGGCCAACCCGACTCTCTCCACGCTGTACAGCCT 1542
DB 204 GAGCCAGTACCGACATCGCGGACTCACCGCGACTCGGGCATCAGGTCTGCCGGTGTGC 263
QY 1543 CTGCACCGAGGTGAGCAGCAGGAGTTCCTGGGTAAGTGAGGAGACGGCAGAGGGCCCCA 1602
DB 264 GGTGGAGGAGGAGGAGGAGGANNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 323
QY 1603 GTCTGTGACGAGAGGCTTCTAGAAAGATCCCCCTGGTGTGACACACTGATGGGG 1662
DB 324 NNN 383
QY 1663 CAGGGTCTGAGGATAGAGGACCGGGGAGGCGCTCCCATGTCATGGTCATGCAGGTACA 1722
DB 384 NNNNGGGNNNNNNNNNNNNNGNNGNNGNNGNNGNNGNNGNNGNNGGNGG 443


```
LOCUS      BC028169      3328 bp      mRNA      linear      HTC 04-MAR-2003
DEFINITION Homo sapiens, Similar to DKFZF434B103 protein, clone IMAGE:5244820,
mRNA.
ACCESSION  BC028169
VERSION    BC028169.1  GI:24081060
KEYWORDS   HTC.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 3328)
AUTHORS    Strausberg, R.
TITLE      Direct Submission
JOURNAL    Submitted (08-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
            Email: cgapsb-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@hgrl.nih.gov
            Akhter, N., Ayèle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
            Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,
            Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
            Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
            Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
            McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
            Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
            Young, A., Zhang, L.-H. and Green, E.D.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAC Plate: 62 Row: n Column: 13
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomesCan gene prediction
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source           1..3328
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clones="IMAGE:5244820"
                     /tissue_type="Brain, fetal, whole pooled"
                     /clone_lib="NIH_MGC_121"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
ORIGIN
Query Match      3.2%; Score 71.6; DB 3; Length 3328;
Best Local Similarity 63.2%; Pred. No. 8.3e-05;
Matches 110; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
Qy 96 TGAATGCTGGGCATCAATAAGACTTTTTCCTGCCAGCTAGTGTGGTAAATAAGAGA 155
Db 1009 TGAATAATTGATGACCTAGATGGAACACATGCTCTGTGATGTCGCCGATGCTCAGAATGAGA 1068
Qy 156 TGGCGTACCTCCTCTGGACCATCAAGAGGACATCATTTGACTATCAAGCTGACCTACG 215
Db 1069 TCCCTACTTCAATCTGGACCACTCGCGGGGATGTGCTCGAGTGTGCTTCTCTCCANG 1128
Qy 216 ACCAGATGCTGAACCACTAGCAAGACAGCCTCTTTCACCACCAAGGTGAGCC 269
Db 1129 ATCAGATGATAAACCACTACGCCGGGCTGGCTCTCTTTACCAAAAGTGGGTC 1182
RESULT 15
```

```
CNS03OAX/c
LOCUS      1040 bp      DNA      linear      GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            041N21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL253122
VERSION    AL253122.1  GI:7974134
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE  1
AUTHORS    Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
            Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
            Saurin, W. and Weissenbach, J.
TITLE      Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
JOURNAL    Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE    20296633
PUBMED     10835645
REFERENCE  2
AUTHORS    Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
            Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
            Saurin, W., Bernot, A. and Weissenbach, J.
TITLE      Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL    Genome Res. 10 (7), 939-949 (2000)
MEDLINE    20359837
PUBMED     10899143
REFERENCE  3 (bases 1 to 1040)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT    This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source           1..1040
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone_lib="G"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : COBG041CG11SP1-end :
                     PUC-Ori"
ORIGIN
Query Match      3.2%; Score 71.4; DB 9; Length 1040;
Best Local Similarity 46.2%; Pred. No. 8e-05;
Matches 150; Conservative 22; Mismatches 153; Indels 0; Gaps 0;
Qy 1659 GGGCGAGGGTCTCAGAGATAGAGCAGCGGGAGAGGCTCCCATGTGTCATGTCAG 1718
Db 496 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 437
Qy 1719 TACAGAGGCGCAGGGGGCCCGGGAGGGAGGGCGGGCGAGTCAAGGAGTGTGTGTGTGC 1778
Db 436 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 377
Qy 1779 TAGGCTTTGTGACAGCGCGGCGAGAGTGTGGGCAAGCCGGCTCCGTGTCGAGGCCG 1838
Db 376 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 317
Qy 1839 AGGGGTGACGTGCGACGCGCGCAGTCAACAGACACTGACAGGAGAGGAGGACGAGCACT 1898
Db 316 GGGKGTGGGGGGGGRGGGGGGRGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 257
```


This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 20:40:04 ; Search time 409,498 Seconds
(without alignments)
8954.609 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgcctgcagcagccggc.....gcctccacaggccctccc 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	69.8	3.1	7218	1	US-08-232-463-14
C 2	54.4	2.4	670689	4	US-09-949-016-12505
C 3	54.4	2.4	670690	4	US-09-949-016-14207
C 4	52.8	2.4	152331	3	US-09-128-155-16
C 5	52.2	2.3	318	3	US-09-165-264-12
C 6	51.2	2.3	319	3	US-09-165-264-8
C 7	51.2	2.3	320	3	US-09-165-264-13
C 8	51	2.3	320	3	US-09-165-264-7
C 9	51	2.3	2561	4	US-09-616-289-48
C 10	50	2.2	320	3	US-09-165-264-14
C 11	49.6	2.2	320	3	US-09-165-264-11
C 12	49.6	2.2	1926	3	US-09-249-585A-2
C 13	49.6	2.2	1926	4	US-09-410-399-3
C 14	49.6	2.2	2580	3	US-09-050-863-2
C 15	49.6	2.2	2580	3	US-09-359-081-2
C 16	49.6	2.2	2580	3	US-09-130-114-1
C 17	49.6	2.2	8705	2	US-09-647-344A-14
C 18	49.6	2.2	9600	3	US-08-910-647-1
C 19	49.6	2.2	9600	3	US-08-910-647-1
C 20	49.6	2.2	10596	1	US-07-884-811-15
C 21	49.6	2.2	10596	1	US-07-885-971-15
C 22	49.6	2.2	10596	1	US-08-087-783A-15
C 23	49.6	2.2	10596	1	US-08-194-088B-15
C 24	49.6	2.2	10596	1	US-08-194-088B-15
C 25	49.6	2.2	10596	5	PCT-US93-04648-15
C 26	49.6	2.2	16080	4	US-09-724-566A-48
C 27	49.6	2.2	16080	4	US-09-471-669A-48

C 28	48.6	2.2	1247	1	US-08-278-729A-32	Sequence 32, Appl
C 29	48.6	2.2	1247	1	US-08-155-343A-32	Sequence 32, Appl
C 30	48.6	2.2	1247	1	US-08-406-672-32	Sequence 32, Appl
C 31	48.6	2.2	1247	1	US-08-643-563A-32	Sequence 32, Appl
C 32	48.6	2.2	1247	1	US-08-643-763A-32	Sequence 32, Appl
C 33	48.6	2.2	1247	1	US-08-462-623-32	Sequence 32, Appl
C 34	48.6	2.2	1247	1	US-08-451-953A-32	Sequence 32, Appl
C 35	48.6	2.2	1247	2	US-08-445-468A-32	Sequence 32, Appl
C 36	48.6	2.2	1247	2	US-08-461-397A-32	Sequence 32, Appl
C 37	48.6	2.2	1247	2	US-08-912-088-32	Sequence 32, Appl
C 38	48.6	2.2	1247	3	US-08-278-730A-32	Sequence 32, Appl
C 39	48.6	2.2	1247	3	US-08-445-467-32	Sequence 32, Appl
C 40	48.6	2.2	1247	3	US-08-480-515A-32	Sequence 32, Appl
C 41	48.6	2.2	1247	3	US-09-170-936-32	Sequence 32, Appl
C 42	48.6	2.2	1247	3	US-08-461-113-32	Sequence 32, Appl
C 43	48.6	2.2	1247	4	US-08-456-033-32	Sequence 32, Appl
C 44	48.6	2.2	1247	4	US-08-643-321-30	Sequence 30, Appl
C 45	48.6	2.2	1247	4	US-08-260-675-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHREIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F18
US-08-232-463-14

Query Match

3.1%; Score 69.8; DB 1; Length 7218;

[illegible]

```

RESULT 2
US-09-949-016-12505
; Sequence 12505, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12505
; LENGTH: 670689
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(670689)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12505

```

Qy	1841	GGGTGACGTGAGACGGCCGAGTCTCAGAGACACTGTGAGGAGAGAGCGCAGCAGACTGG	1900
Db	40279	GGGTGCGGGCGGGGGCGGGGGTGGGGTGGGGCGTCTCGCGGGGGCCCGGGGGGGCGCGCGGG	40338
Qy	1901	GGGTGTGGGAGCAGGCGCTCGGGCCCATCGGCTATCAGAGCGGGAAGTGGGCGGGTTGGG	1960
Db	40339	GCCTGGCGGCGCGCGGGGGCGGGGGCGGTCTTCGGGGGCCCCCGGGGGGGGG	40398
Qy	1961	GAGCCTCCGTGG	1972
Db	40399	CGGTCGCCGGG	40410

RESULT 3
 US-09-949-016-14207
 ; Sequence 14207, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14207
 ; LENGTH: 670690
 ; TYPE: DNA
 ; ORGANISM: Human
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(670690)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-949-016-14207

[illegible]

; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY					
; TITLE OF INVENTION: AND USES THEREOF					
; FILE REFERENCE: 09404/052001					
; CURRENT APPLICATION NUMBER: US/09/128,155					
; CURRENT FILING DATE: 1998-08-03					
; EARLIER APPLICATION NUMBER: US 60/091,650					
; EARLIER FILING DATE: 1998-07-02					
; EARLIER APPLICATION NUMBER: US 60/054,646					
; EARLIER FILING DATE: 1997-08-04					
; NUMBER OF SEQ ID NOS: 18					
; SOFTWARE: FastSeq for Windows Version 3.0					
; SEQ ID NO 16					
; LENGTH: 152331					
; TYPE: DNA					
; ORGANISM: Homo sapiens					
; FEATURE:					
; NAME/KEY: misc feature					
; LOCATION: (1)..(152331)					
; OTHER INFORMATION: n = A,T,C or G					
US-09-128-155-16					
Query Match 2.4%; Score 52.8; DB 3; Length 152331;					
Best Local Similarity 47.8%; Pred. No. 0.063;					
Matches 164; Conservative 0; Mismatches 178; Indels 1; Gaps 1;					
Qy	1659	GGGCGAGGCTTGAGATAGAGACCGGGGAGAGGCCTCCCATGTGTCATGTGTCATGGCAG	1718		
Dd	22189	GGGGGGGGGGCGGGCGGGCGGGCGGGGGGGTTCGGGGGTGCCGGGCGGGCGGGCGGG	22130		
Qy	1719	TACAGAGGCCAGGGCCCCCGGGAGGAGCGGGGACAGTCAAGAGAGTGTGTGTTCTCTGTG	1777		
Dd	22129	GGGGGGGGGGCTCGGGCTGCGGGGGGGGGGGGGGGGGGGCTCGGGGGTGGGGGTGCGTGTG	22070		
Qy	1778	CTAGGCTTGTGGACAGCGCGGCGAGCAGCTGGGGGCAAGCCCGCTCCGTGTCGAGAGCC	1837		
Dd	22069	GGCCGGGTGTGGGGGTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGG	22010		
Qy	1838	GAGGGGTGCAGCTTGGACGCCCGCAGTCA CAGAGACA CTGCAGGGAGAAGGGCAGCAGAC	1897		
Dd	22009	CGGGGGGGGGGGCGGGCGCGGGGGGGGGGGGGGGGGGGGGTGGGGGGCGCGGGGGCGGG	21950		
Qy	1898	TGGGGGTGTGGAGCAGGAGCCTTGGCCCCCATCGGCTATAGGGGGGGGAAGTGGGGCGGGTT	1957		
Dd	21949	CCGGGGGGGGGGGGGGGGGGGGCGCGGCCCGNCGGGGGGGGGGGGGGGGGGGGGGGGGG	21890		
Qy	1958	GGGGAGCCTCCGTGGGCCCTTGGCCCCATCTCCAGTCCCACAGTCC	2000		
Dd	21889	GGGGGGGGGGGNNGNNNGTNNNACCNCACCCACCAACC	21847		
RESULT 5					
US-09-165-264-12					
; Sequence 12, Application US/09165264					
; Patent No. 6197510					
; GENERAL INFORMATION:					
; APPLICANT: Vinayagamoorthy, Thuraiayah					
; TITLE OF INVENTION: Multi-Loci Genomic Analysis					
; FILE REFERENCE: 44747					
; CURRENT APPLICATION NUMBER: US/09/165,264					
; CURRENT FILING DATE: 1998-10-01					
; NUMBER OF SEQ ID NOS: 14					
; SOFTWARE: PatentIn Ver. 2.1					
; SEQ ID NO 12					
; LENGTH: 318					
; TYPE: DNA					
; ORGANISM: Artificial Sequence					
; FEATURE:					
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence					
US-09-165-264-12					
Query Match 2.3%; Score 52.2; DB 3; Length 318;					
Best Local Similarity 48.2%; Pred. No. 0.013;					
Matches 147; Conservative 0; Mismatches 158; Indels 0; Gaps 0;					

[illegible]

QY 1841 GGGTGACGCTGGACGCGCGCAGTGCACAGACACATGTCAGGAGAGAGGGCAGGCAGACTGG 1900
DB 945 GCGGGGGGGCTGTGCTGTCTGTGTGCGCGCGCGCGCGCTGTGCGCGCGCGCGAGCG 886
QY 1901 GGGTGTGGAGCAGGCTGGGGCCCCATCGGCTATGAGGCGGGAAAGTGGGGCGGGTTGG 1960
DB 885 GCGACTCCCGGCGCGACG 826
QY 1961 GAGCTTCGTGCGCTGGCGCCCATCTCCAGTCCCAAGTCTGCTCGACAGATAGGGGG 2020
DB 825 GCGCGCGCGCGCGGAGCT 766
QY 2021 AGCGTGTGCTCTTTCAGAGACTTCGCGCGCGCACCATGGCATCCAGCATCTCAAGTG 2080
DB 765 GCGCGCGGGGGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 706
QY 2081 GGTGTGTCAGCCACAGAGCTGCAGCAGGAGCAGCAGAGCAAGCCAGGACACAGAGGA 2140
DB 705 CG 646
QY 2141 GAGGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGGTGGCTGGCGCGCGCGCGCGAGT 2200
DB 645 GCGCGCGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 586
QY 2201 GCAGGCGCCACAGAGCTCGGGGCGCTCCACAGGGCG 2235
DB 585 GCGCGCGCGCGCGCGGGGTGGTCCGCGCGCGGGC 551

RESULT 10
US-09-165-264-14
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14

Query Match 2.2%; Score 50; DB 3; Length 320;
Best Local Similarity 47.7%; Pred. No. 0.042;
Matches 146; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 1660 GGGCAGGCTCTGAGGATAGAGGACCGGGAGAGGCTCCCATGTCATGGTTCATGGCAGT 1719
DB 1 GGG 60
QY 1720 ACAGAGGCGCAGGGCCCCCGGAGGAGGCGCGCAGTCAAGAGAGTGTGTGTTCTGTGCT 1779
DB 61 GGG 120
QY 1780 AGGCTTGTGACAGCGCGCGCAGAGCTGGGCAAGCGCGCTCCGTGTCGAGGCCGA 1839
DB 121 GGG 180
QY 1840 GGGGTGCACTGACGCGCGCAGTGCACAGACACTGCAGGAGAGAGGCGAGCAGACTG 1899
DB 181 GGG 240
QY 1900 GGGGTGTGGAGCAGGCTGGCGCCCATCGGCTATGAGCGCGGGAAGTGGGCGCGGTTGG 1959
DB 241 GGG 300

QY 1960 GGAGCC 1965
DB 301 TGTGCC 306
RESULT 11
US-09-165-264-11
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

Query Match 2.2%; Score 49.6; DB 3; Length 320;
Best Local Similarity 47.7%; Pred. No. 0.052;
Matches 145; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 1660 GGGCAGGCTCTGAGGATAGAGGACCGGGAGAGGCTCCCATGTCATGGTTCATGGCAGT 1719
DB 1 GGG 60
QY 1720 ACAGAGGCGCAGGGCCCCCGGAGGAGGCGCGCAGTCAAGAGAGTGTGTGTTCTGTGCT 1779
DB 61 GGG 120
QY 1780 AGGCTTGTGACAGCGCGCGCAGCAGCTGGGCAAGCGCGCTCCGTGTCGAGGCCGA 1839
DB 121 GGG 180
QY 1840 GGGGTGCACTGACGCGCGCAGTGCACAGACACTGCAGGAGAGAGGCGAGCAGACTG 1899
DB 181 GGG 240
QY 1900 GGGGTGTGGAGCAGGCTGGCGCCCATCGGCTATGAGCGCGGGAAGTGGGCGCGGTTGG 1959
DB 241 GGG 300

RESULT 12
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)

; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Query Match	2.2%;	Score 49.6;	DB 3;	Length 1926;
Best Local Similarity	45.7%;	Pred. No. 0.09;		
Matches 210;	Conservative 0;	Mismatches 249;	Indels 1;	Gaps 1;
Qy	1737	CGGGAGGAGGGGGGGCAGTCAAGAGTGTGTGGTTCTGTCTAGGCTTGTGGACAGCGC	1796	
Db	476	CAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGACGAGG	535	
Qy	1797	GGCGAGCAGCTGGGGCAAG-GCCGGCTTCGTGTGTCGAGGCCGAGGGGTGCAGCTGCAAGC	1855	
Db	536	GAGGGGCAGGAGGGGCAGGACAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGACGAG	595	
Qy	1856	GCCGCAGTCAAGACACTGCAAGGAGAAAGGGCAGGCAGACACTGGGGGTGTGGAGCAGG	1915	
Db	596	GAGGAGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAG	655	
Qy	1916	CCTGGGGCCCATCGGCTATGAGGGCGGGAAGTGGGGCGGGTTGGGGAGCCTTCGTGGCCC	1975	
Db	656	GAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAG	715	
Qy	1976	TGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGCGAGGCTGTGTGTCT	2035	
Db	716	GGGCAGGAGCAGGAGGGGCAGGACAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAG	775	
Qy	2036	TTCAGAAAGATTTCGGGCGCACCATGGCATTCAGCATTCCTCAAGTGGGTGGTTCAGCCACCA	2095	
Db	776	GAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAG	835	
Qy	2096	GAGCTGCAGCAGGAGCAGCAGAACAGCCAGGGACACAGAGGAGGAGGCCCGGGAGCAG	2155	
Db	836	GAGCAGGAGGAGGGGCAGAGGGGCAGGACAGGAGGGGCAGGAGGGGCAGGAGCAGGAG	895	
Qy	2156	CGACCTGAGCAGCAGGCAAGGTGCGCTGGGCCGGGGCAGG	2195	
Db	896	GGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG	935	

```

RESULT 13
US-09-410-399-3
; Sequence 3, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein-Barr virus
US-09-410-399-3

```

	Query Match	2.2%;	Score 49.6;	DB 4;	Length 1926;
	Best Local Similarity	45.7%;	Pred. No. 0.09;		
	Matches 210; Conservative	0;	Mismatches 249;	Indels 1;	Gaps 1;
QY	1737	CGGAGGGGGCGGCAGCTCAGGAGTGTGTGTTCTTGCTAGGCTTTGTGNACAGCG			1796
Db	476	CAGGAGGGGGCAGAGGGGGCAGGAGGGGCGAGGAGGAGGGCGCAGGAGCAGGAG			535
QY	1797	GGCGAGCAGCTGGGGCAAG - GCCGGCTCCGTGGTCGAGGCCGAGGGGGTGCAGCTGGACG			1855
Db	536	GAGGGCGAGGGGGCAGGACGAGGAGGGGCGAGGAGGGGGCAGGAGGGCAGGAACAG			595
QY	1856	GC CGCAGTACAGAGACACTGCAGGGGAGAAGGGCAGGCAGACTGGGGGGTGTGGGAGCAGG			1915

Db	596	GAGGAGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG	655
Qy	1916	CCTGGGCCCATCGGCTATGAGGCGGGAGTGGGGCGGGTTGGGGAGCCTCCGTGGCCC	1975
Db	656	GAGCAGGAGGAGGGCAGGAGGGGCAGGAGGGCAGGAGCAGGAGGGCAGGAGCAGGAG	715
Qy	1976	TGGCCCATCTCAGTCCCATGCTCTGGCTTCGACAGATAGGGCGAGGCTGTGCTGTCTCT	2035
Db	716	GGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAG	775
Qy	2036	TTCAGAGACTTCCGGCGCACCATGGCATCCAGCATCTCACTGGGTGGTCAGCCACCA	2095
Db	776	GAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAG	835
Qy	2096	GAGCTGCAGCAGGAGCAGAGAGCAGCCACGGGACACAGAGGAGGAGGAGCCGGGAGCAG	2155
Db	836	GAGCAGGAGAGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAG	895
Qy	2156	CGACTCTAGCAGCAGGCAGGTGCTTGGGCCCGGGCAGG	2195
Db	896	GGCAGGAGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG	935

```

RESULT 14
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; TITLE OF INVENTION: System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,863
; FILING DATE: 30-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-09-050-863-2

```

Query Match 2.2%; Score 49.6; DB 3; Length 2580;
Best Local Similarity 45.7%; Pred. No. 0.099;
Matches 210; Conservative 0; Mismatches 249; Indels 1; Gaps 1;
Qy 1737 CCGGAGGGAGGGCGGCAGCTCAGGAGTGTGTGTTCTGTCTAGGCTTGTGGACAGCCG 1796

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 23:48:14 ; Search time 1728.82 Seconds
(without alignment)
8666.982 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgcctgcagcagccggc.....gcctccacaggggccctccc 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079536 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US10J_NEW_PUB.seq.*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2241	100.0	2241	19	US-10-615-659-27
2	2241	100.0	2241	19	US-10-635-977-27
3	2241	100.0	101270	20	US-10-723-860-631
4	1795.8	80.1	3465	19	US-10-615-659-11
5	1795.8	80.1	3465	19	US-10-635-977-11
6	1742.8	77.8	3554	19	US-10-615-659-12
7	1742.8	77.8	3554	19	US-10-635-977-12

8	272.4	12.2	1838	19	US-10-615-659-1	Sequence 1, Appli
9	272.4	12.2	1838	19	US-10-635-977-1	Sequence 1, Appli
10	272.4	12.2	1939	19	US-10-615-659-9	Sequence 9, Appli
11	272.4	12.2	1939	19	US-10-635-977-9	Sequence 9, Appli
12	138.4	6.2	1859	19	US-10-615-659-10	Sequence 10, Appli
13	138.4	6.2	1859	19	US-10-635-977-10	Sequence 10, Appli
14	69.2	3.1	452	9	US-09-728-446-1073	Sequence 1073, Ap
15	67.8	3.0	1042	22	US-10-450-763-8232	Sequence 8232, Ap
16	67.8	3.0	2250	22	US-10-450-763-5187	Sequence 5187, Ap
c 17	65	2.9	778	18	US-10-424-599-54839	Sequence 54839, A
c 18	63.2	2.8	1092	19	US-10-767-701-9739	Sequence 9739, Ap
19	62.6	2.8	740	20	US-10-425-115-131885	Sequence 131885,
20	60.8	2.7	5452	15	US-10-017-161-1481	Sequence 1481, Ap
21	60.8	2.7	5452	17	US-10-292-798-1189	Sequence 1189, Ap
c 22	60	2.7	822	20	US-10-425-115-102047	Sequence 102047, A
c 23	59.6	2.7	630	19	US-10-437-963-18774	Sequence 18774, A
c 24	59.4	2.7	891	20	US-10-425-115-146045	Sequence 146045,
c 25	58.8	2.6	766	20	US-10-425-115-82176	Sequence 82176, A
c 26	58.2	2.6	629	19	US-10-021-323-9375	Sequence 9375, Ap
27	57.4	2.6	588	19	US-10-021-323-9377	Sequence 9377, Ap
c 28	57.4	2.6	925	19	US-10-437-963-44536	Sequence 44536, A
c 29	57	2.5	1327	17	US-10-398-221-1265	Sequence 1265, Ap
c 30	56.8	2.5	700	20	US-10-425-115-14077	Sequence 14077, A
c 31	56.8	2.5	3163	15	US-10-017-161-1857	Sequence 1857, Ap
32	56.8	2.5	3163	17	US-10-292-798-1513	Sequence 1513, Ap
c 33	56.4	2.5	673	20	US-10-425-115-11145	Sequence 11145, A
34	56	2.5	12733	14	US-10-032-393-47	Sequence 47, Appli
c 35	56	2.5	12739	14	US-10-032-393-8	Sequence 8, Appli
c 36	55.4	2.5	959	19	US-10-437-963-48694	Sequence 48694, A
37	55.4	2.5	985	20	US-10-425-115-91623	Sequence 91623, A
38	55	2.5	3133	15	US-10-017-161-1483	Sequence 1483, Ap
39	55	2.5	3133	17	US-10-292-798-1191	Sequence 1191, Ap
c 40	54.8	2.4	497	18	US-10-424-599-133788	Sequence 133788,
c 41	54.6	2.4	514	18	US-10-424-599-104895	Sequence 104895,
c 42	54.4	2.4	815	18	US-10-424-599-20495	Sequence 20495, A
43	54.2	2.4	931	14	US-10-198-846-8585	Sequence 8585, Ap
44	54	2.4	771	18	US-10-424-599-107958	Sequence 107958, A
45	53.8	2.4	600	20	US-10-363-345A-39087	Sequence 39087, A

ALIGNMENTS

RESULT 1
US-10-615-659-27
; Sequence 27, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-27

Query Match	100.0%	Score 2241;	DB 19;	Length 2241;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2241;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TGCTGCAGCAGACCCGGCTGGGATCCTACAGTCGACGCATTTGCTCTGATGTT	60	
Db	1	TGCTGCAGCAGACCCGGCTGGGATCCTACAGTCGACGCATTTGCTCTGATGTT	60	
Qy	61	GCCTCACTCCATCCATCCTGGCTGGTGTGAACCTTGATGCTGGGCATCAATAAGACT	120	


```
QY 1261 ATGCTCCACCGCTGGAGGGGTGTTCTGGGGCCCTGGGGACCGGTATATCTGGGGAGG 1320
Db |||||
QY 58878 ATGCTCCACCGCTGGAGGGGTGTTCTGGGGCCCTGGGGACCGGTATATCTGGGGAGG 58819
Db |||||
QY 1321 GGGAGGGATCCGCCCTTAGGGGTGCTGTGTGGTGGCCCATAGCAGGCAAGTCTGAGGTTGG 1380
Db |||||
QY 58818 GGGAGGGATCCGCCCTTAGGGGTGCTGTGTGGTGGCCCATAGCAGGCAAGTCTGAGGTTGG 58759
Db |||||
QY 1381 ACGGCCGCTCTCGGCTTGTGAGCGAGGAGAGCTGAGCTGAGCTGAGCTGCCCTCTCGCTAA 1440
Db |||||
QY 58758 ACGGCCGCTCTCGGCTTGTGAGCGAGGAGAGCTGAGCTGAGCTGCCCTCTCGCTAA 58699
Db |||||
QY 1441 CACGCCACTGTCTCGGAGATCGGGCTGTGGTGAACATGCGGAGCCTGCCCTGTAGCT 1500
Db |||||
QY 58698 CACGCCACTGTCTCGGAGATCGGGCTGTGGTGAACATGCGGAGCCTGCCCTGTAGCT 58639
Db |||||
QY 1501 CCGCGCCAAACCCGACTCCTCTTCCACGCTGTACAGCCTCTGCACCGAGGTGAGCA 1560
Db |||||
QY 58638 CCGCGCCAAACCCGACTCCTCTTCCACGCTGTACAGCCTCTGCACCGAGGTGAGCA 58579
Db |||||
QY 1561 GCAGGATTTCTGGGTAGTAGAGAGACGGCAGAGGGCCCAAGTGTGTGACAGAGAGG 1620
Db |||||
QY 58578 GCAGGATTTCTGGGTAGTAGAGAGACGGCAGAGGGCCCAAGTGTGTGACAGAGAGG 58519
Db |||||
QY 1621 CTTCTAGAAAGATCCCTGCTGTGACAGACACTGATGGGGCAGGGTCTGAGGATAGAG 1680
Db |||||
QY 58518 CTTCTAGAAAGATCCCTGCTGTGACAGACACTGATGGGGCAGGGTCTGAGGATAGAG 58459
Db |||||
QY 1681 GACCGGGAGAGGCTCCCATGCTCATGTGATGTCAGTACAGAGCCAGGGGCCCGGG 1740
Db |||||
QY 58458 GACCGGGAGAGGCTCCCATGCTCATGTGATGTCAGTACAGAGCCAGGGGCCCGGG 58399
Db |||||
QY 1741 AGGAGGGCGGGCAGTCAAGGAGTGTGTGCTGTGCTAGGCTTTGTGACAGCGCGCG 1800
Db |||||
QY 58398 AGGAGGGCGGGCAGTCAAGGAGTGTGTGCTGTGCTAGGCTTTGTGACAGCGCGCG 58339
Db |||||
QY 1801 AGCAGCTGGGGCAAGCCCGGCTCCGTGTCGAGGCCGAGGGGTGACGTGAGCGGGCCG 1860
Db |||||
QY 58338 AGCAGCTGGGGCAAGCCCGGCTCCGTGTCGAGGCCGAGGGGTGACGTGAGCGGGCCG 58279
Db |||||
QY 1861 AGTCACAGACACTGTCAGGGAGAGGGCAGGCAGACTGGGGGTGTGGGAGCAGGCTGG 1920
Db |||||
QY 58278 AGTCACAGACACTGTCAGGGAGAGGGCAGGCAGACTGGGGGTGTGGGAGCAGGCTGG 58219
Db |||||
QY 1921 GCGCCATCGGCTATCAGGGCGGGAAGTGGGGCGGTGGGGAGCCTCCGTGGCCCTGGCC 1980
Db |||||
QY 58218 GCGCCATCGGCTATCAGGGCGGGAAGTGGGGCGGTGGGGAGCCTCCGTGGCCCTGGCC 58159
Db |||||
QY 1981 CCATCTCAGTCCCAGTCTCGGCTCGGACAGATAGGGGAGGCTGTGCTGCTTTTTCAG 2040
Db |||||
QY 58158 CCATCTCAGTCCCAGTCTCGGCTCGGACAGATAGGGGAGGCTGTGCTGCTTTTTCAG 58099
Db |||||
QY 2041 AAGACTTCGGGCGCACCATGGCATCCAGCATCTCAAGTGGTGGTGTGAGCCACAGAGCT 2100
Db |||||
QY 58098 AAGACTTCGGGCGCACCATGGCATCCAGCATCTCAAGTGGTGGTGTGAGCCACAGAGCT 58039
Db |||||
QY 2101 GCAGCAGGACGACAGAGAGCCAGCCAGGAGGAGGAGGCGGGAGCAGCGGAC 2160
Db |||||
QY 58038 GCAGCAGGACGACAGAGAGCCAGCCAGGAGGAGGAGGCGGGAGCAGCGGAC 57979
Db |||||
QY 2161 TGAGCAGGAGCAAGGTGCGTGGCGCGGGCAGGGCAGTGCAGGCCACACAGAGCTCGG 2220
Db |||||
QY 57978 TGAGCAGGAGCAAGGTGCGTGGCGCGGGCAGGGCAGTGCAGGCCACACAGAGCTCGG 57919
Db |||||
QY 2221 GGCCTCCACAGGGGCCCTCCC 2241
Db |||||
QY 57918 GGCCTCCACAGGGGCCCTCCC 57898
```

RESULT 4

US-10-615-659-11

; Sequence 11, Application US/10615659

; Publication No. US20040157234A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN

; FILE REFERENCE: D0283 NP

; CURRENT APPLICATION NUMBER: US/10/615,659

; PRIOR FILING DATE: 2003-07-09

; PRIOR APPLICATION NUMBER: U.S. 60/394,725

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 11

; LENGTH: 3465

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-10-615-659-11

```
Query Match 80.1%; Score 1795.8; DB 19; Length 3465;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1797; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 381 CCCAGGACTCCTGTGTTCTAAGGCCGTGGCCAGAACTCACTCGGTGCCCCACCCACCCCGCAGC 440
Db |||||
QY 441 ACCCTGTGCCCTTGTCTGTCTGTGTTGTAATCCGGGGCCCCAGAGTCCCTCTCT 500
Db |||||
QY 61 ACCCTGTGCCCTTGTCTGTCTGTGTTGTAATCCGGGGCCCCAGAGTCCCTCTCT 120
Db |||||
QY 501 CAGGGCAGAGGCGCAAGATGGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGA 560
Db |||||
QY 121 CAGGGCAGAGGCGCAAGATGGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGA 180
Db |||||
QY 561 GGGGGCTGTGGGACCCAGAGGGAAGAGGTGCGGTCTCCCGCCAGACACCGGGCACTCCCGCG 620
Db |||||
QY 181 GGGGGCTGTGGGACCCAGAGGGAAGAGGTGCGGTCTCCCGCCAGACACCGGGCACTCCCGCG 240
Db |||||
QY 621 CCTCCCCGGGCTCCAGCCCTGCACAAAGAGCTTGGCCACACCTTCCCAACAGGGCCCA 680
Db |||||
QY 241 CCTCCCCGGGCTCCAGCCCTGCACAAAGAGCTTGGCCACACCTTCCCAACAGGGCCCA 300
Db |||||
QY 681 GGGCTGAGGGCTGCCCGCAGGACGAGGGTGTGGGACCTGTGAGGGAGGGGTCCCGGAA 740
Db |||||
QY 301 GGGCTGAGGGCTGCCCGCAGGACGAGGGTGTGGGACCTGTGAGGGAGGGGTCCCGGAA 360
Db |||||
QY 741 GAGGGGCTTCCCTGGCAGAGGTCCCTCTCAGCAGGCCAGGTGCGGTGCTCAGCACA 800
Db |||||
QY 361 GAGGGGCTTCCCTGGCAGAGGTCCCTCTCAGCAGGCCAGGTGCGGTGCTCAGCACA 420
Db |||||
QY 801 GTGTGGGGCGGAGTGCAGGACAAAGTTCCCTTCCGCACTTAATACCCAGGGTCAGGCC 860
Db |||||
QY 421 GTGTGGGGCGGAGTGCAGGACAAAGTTCCCTTCCGCACTTAATACCCAGGGTCAGGCC 480
Db |||||
QY 861 AGCGCCAGCGCTGTAGTGGCAACATGGCCCCCTTCAAGACCCCGTGTGAGAGCCAGTC 920
Db |||||
QY 481 AGCGCCAGCGCTGTAGTGGCAACATGGCCCCCTTCAAGACCCCGTGTGAGAGCCAGTC 540
Db |||||
QY 921 AGCGTGTGCTGTGTTTCTGGGGCCAGCCACTTTTCTTAAACAGGGTGCACAAACAG 980
Db |||||
QY 541 AGCGTGTGCTGTGTTTCTGGGGCCAGCCACTTTTCTTAAACAGGGTGCACAAACAG 600
Db |||||
QY 981 ACATCGGGGTGCGGGGACTTTCACGATATGCTCTCTAATGCGCGGTGTACTCTGCGGTGT 1040
Db |||||
QY 601 ACATCGGGGTGCGGGGACTTTCACGATATGCTCTCTAATGCGCGGTGTACTCTGCGGTGT 660
Db |||||
QY 1041 GGTGGGAGTGTGCGGATGTGGCTTCCGGGCCCTTCCAGCGGCTCTCTCTCTCTCTCT 1100
Db |||||
QY 661 GGTGGGAGTGTGCGGATGTGGCTTCCGGGCCCTTCCAGCGGCTCTCTCTCTCTCTCTCT 720
Db |||||
QY 1101 GGGGTCTGAGGTGGGACACCCCTGTGGCCCGCTTTGGGGGCTTGGAGGAGCCCCAGTGC 1160
Db |||||
QY 721 GGGGTCTGAGGTGGGACACCCCTGTGGCCCGCTTTGGGGGCTTGGAGGAGCCCCAGTGC 780
```

QY 1161 CCACGCCCTCGGTGCCCTCAGCTGCAGCCTCCCAATTCCTGGGGCTGCCACGCCCTC 1220
DB CCACGCCCTCGGTGCCCTCAGCTGCAGCCTCCCAATTCCTGGGGCTGCCACGCCCTC 840
QY 1221 CACTGCTCAAGGATGCGTCTGAGCTGGCTGCACATCCATGCTCCACCGGTGGAGG 1280
DB CACTGCTCAAGGATGCGTCTGAGCTGGCTGCACATCCATGCTCCACCGGTGGAGG 900
QY 1281 GGGTGTTCGGGGCCCTGGGGGACCGTATATCTGGGAGGGGAGGATCCGCCCTAGGG 1340
DB GGGTGTTCGGGGCCCTGGGGGACCGTATATCTGGGAGGGGAGGATCCGCCCTAGGG 960
QY 1341 GTGCTGTGTGTCGCCATACGCCAAGCAAGTCTGAGGTGGACGCCCGTCTCCGCTTGT 1400
DB GTGCTGTGTGTCGCCATACGCCAAGCAAGTCTGAGGTGGACGCCCGTCTCCGCTTGT 1020
QY 1401 GAGCGAGCAGAGCTGAGCTGAGCTGCCCTCTGCTTAAACACGCCACTGTCTTGAGA 1460
DB GAGCGAGCAGAGCTGAGCTGAGCTGCCCTCTGCTTAAACACGCCACTGTCTTGAGA 1080
QY 1461 TCGGGCTGTGCTGAACATCGGAGCCTGCCCTGTGCTAGCTCCCGGCCAACCCGACTCCT 1520
DB TCGGGCTGTGCTGAACATCGGAGCCTGCCCTGTGCTAGCTCCCGGCCAACCCGACTCCT 1140
QY 1521 TCTTCCCAAGCTGTACAGCCTCTGCACCGAGGTGAGCAGCAGGAGTTCCTGGGTAAGT 1580
DB TCTTCCCAAGCTGTACAGCCTCTGCACCGAGGTGAGCAGCAGGAGTTCCTGGGTAAGT 1200
QY 1581 GAGGAGCGCAGAGGGCCCCAGTGTCTGCTCAGCAGAGGCTTCTAGAAAGATCCCCCTG 1640
DB GAGGAGCGCAGAGGGCCCCAGTGTCTGCTCAGCAGAGGCTTCTAGAAAGATCCCCCTG 1260
QY 1641 GTGCTGAGACAGACTGATGGGGCAGGGTCTCAGGATAGAGACCGGGAGAGGCTCCCA 1700
DB GTGCTGAGACAGACTGATGGGGCAGGGTCTCAGGATAGAGACCGGGAGAGGCTCCCA 1320
QY 1701 TGGTCATGGTCATGCGAGTACAGAGGCCAGGGCCCCGGGAGGAGGGCGGGCAGTCAAG 1760
DB TGGTCATGGTCATGCGAGTACAGAGGCCAGGGCCCCGGGAGGAGGGCGGGCAGTCAAG 1380
QY 1761 GAGTGTGTGGTCTGTGCTAGGCTTGTGGACAGCGCGCGAGCAGCTGGGGCAGGGCCG 1820
DB GAGTGTGTGGTCTGTGCTAGGCTTGTGGACAGCGCGCGAGCAGCTGGGGCAGGGCCG 1440
QY 1821 CTCGCTGTGTCGAGGCCGAGGGGTGCAGCTGCGACGCGCGCAGTCAACAGACACTGCAGG 1880
DB CTCGCTGTGTCGAGGCCGAGGGGTGCAGCTGCGACGCGCGCAGTCAACAGACACTGCAGG 1500
QY 1881 GAGAAAGGACAGCAGACTGGGGGTGTGGGAGCAGGCCCTGGGGCCCCCATCGGCTATGAGGGC 1940
DB GAGAAAGGACAGCAGACTGGGGGTGTGGGAGCAGGCCCTGGGGCCCCCATCGGCTATGAGGGC 1560
QY 1941 GGGAAAGTGGGCGGGTGGGAGAGCTCGTGCGCCCTGGCCCCCATCTCCAGTCCCCAGTCC 2000
DB GGGAAAGTGGGCGGGTGGGAGAGCTCGTGCGCCCTGGCCCCCATCTCCAGTCCCCAGTCC 1620
QY 2001 TGGCTCGGACAGATAGGCGAGGCTGTGCTGCTCTTTTCAAGAACTTCCCGCGCACCATG 2060
DB TGGCTCGGACAGATAGGCGAGGCTGTGCTGCTCTTTTCAAGAACTTCCCGCGCACCATG 1680
QY 2061 GCATCCAGCATCCTCAAGTGGGTGGTTCAGCCACCAGAGCTGCAGCAGGAGCAGCAGAAGC 2120
DB GCATCCAGCATCCTCAAGTGGGTGGTTCAGCCACCAGAGCTGCAGCAGGAGCAGCAGAAGC 1740
QY 2121 AAGCCAGGGACAGAGGAGGCGGGAGAGCAGGACCTGAGCAGCAGCAGGAGTGC 2179
DB AAGCCAGGGACAGAGGAGGCGGGAGAGCAGGACCTGAGCAGCAGCAGGAGTGC 1799

RESULT 5
US-10-635-977-11
; Sequence 11, Application US/10635977

; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-11

Query Match 80.1%; Score 1795.8; DB 19; Length 3465;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1797; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 381 CCACGAGTCTCTGTCTTAAGGCCGTGGCCAGAAATCACTCGGTGCCACACCCACCCGAGC 440
DB 1 CCACGAGTCTCTGTCTTAAGGCCGTGGCCAGAAATCACTCGGTGCCACACCCACCCGAGC 60
QY 441 ACCCTGTGCCCTTTGTCTGTCTGTGGTGAATCCGGGGCCCAAGAGTCCCTCTCT 500
DB 61 ACCCTGTGCCCTTTGTCTGTCTGTGGTGAATCCGGGGCCCAAGAGTCCCTCTCT 120
QY 501 CAGGGCACAGAGGCCAAAGATGGGGCTGACTGGGGGTGCCACCGGGCTTTGGGTGCTGA 560
DB 121 CAGGGCACAGAGGCCAAAGATGGGGCTGACTGGGGGTGCCACCGGGCTTTGGGTGCTGA 180
QY 561 GGGGGCTGTGGGACCCAGGGGGAAGAGTGGCGTCCCCCAGGACACGGGGACTCCCCG 620
DB 181 GGGGGCTGTGGGACCCAGGGGGAAGAGTGGCGTCCCCCAGGACACGGGGACTCCCCG 240
QY 621 CTTCCCCCGGCTCCAGCCCTGCACAAAGCAGCTTGCACACCTTCCACCAAGGGCCA 680
DB 241 CTTCCCCCGGCTCCAGCCCTGCACAAAGCAGCTTGCACACCTTCCACCAAGGGCCA 300
QY 681 GGGCTGAGGGCTGCCACAGGAGCGAGGGTGTGGGACCTCTGCTGAGGAGGGGTCCCGAA 740
DB 301 GGGCTGAGGGCTGCCACAGGAGCGAGGGTGTGGGACCTCTGCTGAGGAGGGGTCCCGAA 360
QY 741 GAGGGGTTCTCTGGGCAAGAGTCTCTCAGCAGGCCAGGTGGCGCTGCTCAGCACA 800
DB 361 GAGGGGTTCTCTGGGCAAGAGTCTCTCAGCAGGCCAGGTGGCGCTGCTCAGCACA 420
QY 801 GTGTGGGGCGAGGTGCAGGACCAAGTTCCTTCCGACCTTAATACCCAGGGTCAGGCC 860
DB 421 GTGTGGGGCGAGGTGCAGGACCAAGTTCCTTCCGACCTTAATACCCAGGGTCAGGCC 480
QY 861 AGGCCACAGCGCTGTAGTGGCAACATGGCCCCCTTCAAAGACCCCGTGTGCAGAGCCAGTC 920
DB 481 AGGCCACAGCGCTGTAGTGGCAACATGGCCCCCTTCAAAGACCCCGTGTGCAGAGCCAGTC 540
QY 921 AGCGTGGCTGTGTTTTCTGGGGCCAGCACAATTTTTTCTTAAACAGGGTGAACAAACAG 980
DB 541 AGCGTGGCTGTGTTTTCTGGGGCCAGCACAATTTTTTCTTAAACAGGGTGAACAAACAG 600
QY 981 ACATCGGGGTCCGGGACTTTCAGATATGCTCTCTAATGGCGGTCTACTCTCGCTGT 1040
DB 601 ACATCGGGGTCCGGGACTTTCAGATATGCTCTCTAATGGCGGTCTACTCTCGCTGT 660
QY 1041 GGTGGGAGTGTGGGATGTGGGTTCGGGGCCCTGCGAGCCGTCTCTCCCTCTCTCT 1100
DB 661 GGTGGGAGTGTGGGATGTGGGTTCGGGGCCCTGCGAGCCGTCTCTCCCTCTCTCT 720
QY 1101 GGGGTCTGAGGTGTGGGACACCCCTGTGGCCCGCTTTGGGGGCTGGAGGGAGCCCCAGTGC 1160

```

Db 721 GGGGCTGAGGTGTGGGACACCTGTGGCCCGTTTGGGGCCCTGGAGGAGCCCAAGTGC 780
QY 1161 CCACGCCCTCGGTGCGCTCAGCTGAGCCTTCCCAATTCCTTGAGGCTTGCACGCGCTTC 1220
Db 781 CCACGCCCTCGGTGCGCTCAGCTGAGCCTTCCCAATTCCTTGAGGCTTGCACGCGCTTC 840
QY 1221 CACTGCTCAAGGATGCGTCTGAGCTGAGCCTGCGCTGCAATCATCTCCACCGGCTGGAG 1280
Db 841 CACTGCTCAAGGATGCGTCTGAGCTGAGCCTGCGCTGCAATCATCTCCACCGGCTGGAG 900
QY 1281 GGGTGTTCCTGGGGCCCTGGGGACCGTATATCTGGGGAGGGGAGGATTCGCCCTTAGG 1340
Db 901 GGGTGTTCCTGGGGCCCTGGGGACCGTATATCTGGGGAGGGGAGGATTCGCCCTTAGG 960
QY 1341 GTGCTGTGTGTCCTCCCATAGCCAGCAAGTCTGAGGTGGAGCGGCCGCTCTCGCTGTGT 1400
Db 961 GTGCTGTGTGTCCTCCCATAGCCAGCAAGTCTGAGGTGGAGCGGCCGCTCTCGCTGTGT 1020
QY 1401 GAGCGAGGCGAGAGCTGAGCTGAGCTGCCCTCTCTGCTTAACACGCCACTGTCTCTGGAGA 1460
Db 1021 GAGCGAGGCGAGAGCTGAGCTGAGCTGCCCTCTCTGCTTAACACGCCACTGTCTCTGGAGA 1080
QY 1461 TCGGGCTGTGCTGAACATGCGGAGCGCTGCCCTGTGTACGTCCCGGCCAACCCGCTCTCT 1520
Db 1081 TCGGGCTGTGCTGAACATGCGGAGCGCTGCCCTGTGTACGTCCCGGCCAACCCGCTCTCT 1140
QY 1521 TCTTCCACGCTGTACAGCCTCTGACCGAGGTGAGCAGCAGGAGTTCCTGGGTAAAGT 1580
Db 1141 TCTTCCACGCTGTACAGCCTCTGACCGAGGTGAGCAGCAGGAGTTCCTGGGTAAAGT 1200
QY 1581 GAGGAGCGGCGAGAGGCGCCAGTGTCTGACAGAGGCTTCTAGAAGATCCCCCTG 1640
Db 1201 GAGGAGCGGCGAGAGGCGCCAGTGTCTGACAGAGGCTTCTAGAAGATCCCCCTG 1260
QY 1641 GTGCTGAGACAGACTGATGGGCGAGGCTGTAGGATAGAGGACCGGGAGAGGCTTCCCA 1700
Db 1261 GTGCTGAGACAGACTGATGGGCGAGGCTGTAGGATAGAGGACCGGGAGAGGCTTCCCA 1320
QY 1701 TGGTCAATGTCATGTCAGTACAGAGCCAGGGCCCCCGGAGGGAGGGCGGCGACTCAAG 1760
Db 1321 TGGTCAATGTCATGTCAGTACAGAGCCAGGGCCCCCGGAGGGAGGGCGGCGACTCAAG 1380
QY 1761 GAGTGTGTGTTCTGTCTAGCTTGTGGAAGCGGGGAGCAGCTGCGGGCAAGCGCG 1820
Db 1381 GAGTGTGTGTTCTGTCTAGCTTGTGGAAGCGGGGAGCAGCTGCGGGCAAGCGCG 1440
QY 1821 CTCCTGTGTGAGGCGGAGGGGTGAGCTGGACGCGCGCAGTCAAGAGACACTGCAAG 1880
Db 1441 CTCCTGTGTGAGGCGGAGGGGTGAGCTGGACGCGCGCAGTCAAGAGACACTGCAAG 1500
QY 1881 GAGAGGGCGAGCAGACTGGGGGTGTGGAGCAGGCTTGGGCCCCATCGGCTATGAGGGC 1940
Db 1501 GAGAGGGCGAGCAGACTGGGGGTGTGGAGCAGGCTTGGGCCCCATCGGCTATGAGGGC 1560
QY 1941 GGGAGTGGGGCGGGTGGGAGACCTTCGTGGCCCTTGGCCCATCTCCAGTCCCGAGTCC 2000
Db 1561 GGGAGTGGGGCGGGTGGGAGACCTTCGTGGCCCTTGGCCCATCTCCAGTCCCGAGTCC 1620
QY 2001 TGGCTCGGACAGATAGGGCGAGGCTGTGTCTTTTCAAGAGACTTCGGGCGCACCATG 2060
Db 1621 TGGCTCGGACAGATAGGGCGAGGCTGTGTCTTTTCAAGAGACTTCGGGCGCACCATG 1680
QY 2061 GCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGCTGACGAGGAGCAGCAGAGAAGC 2120
Db 1681 GCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGCTGACGAGGAGCAGCAGAGAAGC 1740
QY 2121 AAGCCAGGACAGAGGAGGAGCGCGGAGCAGCAGCTTACGAGCAGGAGGCAAGTGC 2179
Db 1741 AAGCCAGGACAGAGGAGGAGCGCGGAGCAGCAGCTTACGAGCAGGAGGCAAGTGC 1799
```

```

US-10-615-659-12
; Sequence 12, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-12

Query Match 77.8%; Score 1742.8; DB 19; Length 3554;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 2; Indels 30; Gaps 1;

QY 394 GTCTAAGGCGGTGGCCAGAACTACTCGGTGCCCAACCCCAAGCTCCCTCCTCAGGGCAGAGG 453
Db 1 GTCTAAGGCGGTGGCCAGAACTACTCGGTGCCCAACCCCAAGCTCCCTCCTCAGGGCAGAGG 60
QY 454 TTGCTCTGTGTCTCTGGGTGAATCCGGGGCCCAAGAGCTCCCTCCTCAGGGCAGAGG 513
Db 61 TTGCTCTGTGTCTCTGGGTGAATCCGGGGCCCAAGAGCTCCCTCCTCAGGGCAGAGG 120
QY 514 CAAAGATGGGGCTGACTGGGGGTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGGGA 573
Db 121 CAAAGATGGGGCTGACTGGGGGTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGGGA 180
QY 574 CCCAGGGGAGAGGTGCGCTCCCGCCAGACCGGGCACTCCCGCGCTCCCCGGGCT 633
Db 181 CCCAGGGGAGAGGTGCGCTCCCGCCAGACCGGGCACTCCCGCGCTCCCCGGGCT 240
QY 634 CCAGCCCTGCAAAAGCAGCTTGCCACACCTTCCACAAAGGCCAGGGCTGAGGCGCTG 693
Db 241 CCAGCCCTGCAAAAGCAGCTTGCCACACCTTCCACAAAGGCCAGGGCTGAGGCGCTG 300
QY 694 CCCAGGACGAGGTGTGGGACCTCTGTGAGGAGGGGTCCCGAAGAGGGGCTTCCT 753
Db 301 CCCAGGACGAGGTGTGGGACCTCTGTGAGGAGGGGTCCCGAAGAGGGGCTTCCT 360
QY 754 GGCAAGAGGTTCCTCTCAGCAGGCGAGTGGGCTGCTCAGACACAGTGTGGGGGGAG 813
Db 361 GGCAAGAGGTTCCTCTCAGCAGGCGAGTGGGCTGCTCAGACACAGTGTGGGGGGAG 420
QY 814 GTGAGGACAAAGTTCCTCTCCGCACTTAATACCCAGGGTCAGGCGCAGCGCGCTG 873
Db 421 GTGAGGACAAAGTTCCTCTCCGCACTTAATACCCAGGGTCAGGCGCAGCGCGCTG 480
QY 874 CTAGTGGCAACATGGCCCTTCAAAGACCCCGTGTGAGAGCCAGTCAAGTGTGCGCTGTG 933
Db 481 CTAGTGGCAACATGGCCCTTCAAAGACCCCGTGTGAGAGCCAGTCAAGTGTGCGCTGTG 540
QY 934 TTTTCTGGGGCCAGCCACTTTTCTTACAGGGTGACAAACAGACATCGGGGTGCG 993
Db 541 TTTTCTGGGGCCAGCCACTTTTCTTACAGGGTGACAAACAGACATCGGGGTGCG 600
QY 994 GGGACTTCACCATATGCTCTCTAATGGCGGTGCTACTCGCGTGTGGTGAGGAGTGTCT 1053
Db 601 GGGACTTCACCATATGCTCTCTAATGGCGGTGCTACTCGCGTGTGGTGAGGAGTGTCT 660
QY 1054 GGGCATGTGGTGTGCTTCCGGGCCCTGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1113
Db 661 GGGCATGTGGTGTGCTTCCGGGCCCTGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
QY 1114 TGGACACCCCTGTGGCGGCTTTGGGGGCTGTGGAGGAGCCCAAGTGTGCCACCGCCCTCG 1173
```

Db 721 TGGGACACCTGTGGCCCGCTTGGGGGCTCGAGGGAGCCCAAGTGGCCACCGCCCTCGG 780
QY 1174 TGCCCTCAGCTCGACGCTCCCAATTCCTCGGGCTGCCAGCGCCTCCACTCGCTCAAGG 1233
Db 781 TGCCCTCAGCTCGACGCTCCCAATTCCTCGGGCTGCCAGCGCCTCCACTCGCTCAAGG 840
QY 1234 ATGCGTCTGAGCCTGGCCTGCACATCCATGCTCCACCGGCTGGAGGGGGTTCCTGGG 1293
Db 841 ATGCGTCTGAGCCTGGCCTGCACATCCATGCTCCACCGGCTGGAGGGGGTTCCTGGG 900
QY 1294 CCCTGGGGGACCGTATATCTGGGAGGGGGAGGAGATCCGCCCTAGGGGTGCTGTGGTG 1353
Db 901 CCCTGGGGGACCGTATATCTGGGAGGGGGAGGAGATCCGCCCTAGGGGTGCTGTGGTG 960
QY 1354 CCCCATAGCCAGGCAAGTCTCAGGTGGACGGCCGCTCCTCGCTTGTGAGGAGGCGAG 1413
Db 961 CCCCATAGCCAGGCAAGTCTCAGGTGGACGGCCGCTCCTCGCTTGTGAGGAGGCGAG 1020
QY 1414 CTGAGCTGAGCTGCCCTCCTGCTTAAACGCCCACTGTCTCTGGAGATCGGGCTGTGCGT 1473
Db 1021 CTGAGCTGAGCTGCCCTCCTGCTTAAACGCCCACTGTCTCTGGAGATCGGGCTGTGCGT 1080
QY 1474 GAAATGCGGAGCCTGCCCTGGTAGTCCCGGCCAAACCCCGACTCTTCTTCCACGCTG 1533
Db 1081 GAAATGCGGAGCCTGCCCTGGTAGTCCCGGCCAAACCCCGACTCTTCTTCCACGCTG 1140
QY 1534 CTACAGCCTCTGCACCGAGAGTGAGCAGCAGAGTTCCTGGTGAAGTGAGGAGAGCGCAG 1593
Db 1141 CTACAGCCTCTGCACCGAGAGTGAGCAGCAGAGTTCCTGGTGAAGTGAGGAGAGCGCAG 1200
QY 1594 AGGGCCCCAGTGTCTCAGCAGAGAGGCTTCTAGAAAGATCCCCCTGTGTGTGAGACAGA 1653
Db 1201 AGGGCCCCAGTGTCTCAGCAGAGAGGCTTCTAGAAAGATCCCCCTGTGTGTGAGACAGA 1260
QY 1654 CTGATGGGGCAGGGTCTGAGGATAGAGAACCGGGGAGAGGCTCCCATGTGTCATGTCAT 1713
Db 1261 CTGATGGGGCAGGGTCTGAGGATAGAGAACCGGGGAGAGGCTCCCATGTGTCATGTCAT 1320
QY 1714 GGCAGTACAGAGGCGAGGGCCCCGGGAGGAGGCGCGCAGTCAAGGAGTGTGTTTC 1773
Db 1321 GGCAGTACAGAGGCGAGGGCCCCGGGAGGAGGCGCGCAGTCAAGGAGTGTGTTTC 1380
QY 1774 TGTGTAGGCTTGTGGAAGCAGCGCGCAGCAGCTGGGCAAGGCGGCTCCGTGGTGGGA 1833
Db 1381 TGTGTAGGCTTGTGGAAGCAGCGCGCAGCAGCTGGGCAAGGCGGCTCCGTGGTGGGA 1440
QY 1834 GGCAGAGGGTGCAGCTGGAGCGCGCGAGTCAAGAGACACTGCAGGGAGAGGGCAGGC 1893
Db 1441 GGCAGAGGGTGCAGCTGGAGCGCGCGAGTCAAGAGACACTGCAGGGAGAGGGCAGGC 1500
QY 1894 AGACTGGGGGTGGGAGCAGGCTGGGCCCCCATCGGCTATGAGGGCGGAAGTGGGGCG 1953
Db 1501 AGACTGGGGGTGGGAGCAGGCTGGGCCCCCATCGGCTATGAGGGCGGAAGTGGGGCG 1560
QY 1954 GGTGGGGAGCCTCGTGGCCCTGGCCCCCATCTCCAGTCCCGCAGTCTGCTGGTCCGACAGA 2013
Db 1561 GGTGGGGAGCCTCGTGGCCCTGGCCCCCATCTCCAGTCCCGCAGTCTGCTGGTCCGACAGA 1620
QY 2014 TAGGGCAGGCTGTGCTGCTCTTTCA-----GAAG 2043
Db 1621 TAGGGCAGGCTGTGCTGCTCTTTTCAGAGTAGCAGCTGGGAAGAGAGTTCCTTGGAA 1680
QY 2044 ACTTCCGGCGCACCATGCGATCCAGCATCTCAAGTGGGTGTCAGCCACCAGAGCTGCA 2103
Db 1681 ACTTCCGGCGCACCATGCGATCCAGCATCTCAAGTGGGTGTCAGCCACCAGAGCTGCA 1740
QY 2104 GCAGAGCAGCAGAGCAAGCCACGGGACCAAGAGGAGGAGCGGGAGCAGCAGCCTGA 2163
Db 1741 GCAGAGCAGCAGAGCAAGCCACGGGACCAAGAGGAGGAGGCGCGGAGCAGCGACCTGA 1800
QY 2164 GCAGAGCAGAGGTGC 2179
|||||

Db 1801 GCAGAGGCAAGATGC 1816
RESULT 7
; Sequence 12, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-12
Query Match 77.8%; Score 1742.8; DB 19; Length 3554;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 2; Indels 30; Gaps 1;
QY 394 GTCTAAGGCGGTGGCCAGAAATCACTCGGTGCGCCACACCCACCCAGCACCCCTGTGCGCT 453
Db 1 GTCTAAGGCGGTGGCCAGAAATCACTCGGTGCGCCACACCCACCCAGCACCCCTGTGCGCT 60
QY 454 TTGCTCTCTCTCTCTGGTGAATCCGGGGCCCGAGAGCTCCCTCCACAGGCGACAGG 513
Db 61 TTGCTCTCTCTCTCTGGTGAATCCGGGGCCCGAGAGCTCCCTCCACAGGCGACAGG 120
QY 514 CCAAGATGGGCTGACTGGGGCTGCACCGGGCTTTGGGTGTGAGGGGGCTGTGGGA 573
Db 121 CCAAGATGGGCTGACTGGGGCTGCACCGGGCTTTGGGTGTGAGGGGGCTGTGGGA 180
QY 574 CCCAGGGGAAGGTGCCCTCCGCCAGACCGGGCACTCCCGCCGCTCCCGCCGCT 633
Db 181 CCCAGGGGAAGGTGCCCTCCCGCAGCAGCGGGCACTCCCGCCGCTCCCGCCGCT 240
QY 634 CCAGCCCTGCACAAAGCAGCTTGCACACCTTCCCAAGGCCAGGGCCAGGGCTGAGGCTG 693
Db 241 CCAGCCCTGCACAAAGCAGCTTGCACACCTTCCCAAGGGCCAGGGCTGAGGCTG 300
QY 694 CCCAGGACGAGGTGTGGGGAACCTGCTGAGGGAGGGGTCCCGAAAGAGGGGCTTCCCT 753
Db 301 CCCAGGACGAGGTGTGGGGAACCTGCTGAGGGAGGGGTCCCGAAAGAGGGGCTTCCCT 360
QY 754 GGCAACAGAGTCCCTCTCAGCAGGCCAGGTGCGGCTGCCCTCAGCACAGTGTGGGGCGAG 813
Db 361 GGCAACAGAGTCCCTCTCAGCAGGCCAGGTGCGGCTGCCCTCAGCACAGTGTGGGGCGAG 420
QY 814 GTGCAGACAGGTTCCTTCCGACCTAATACCCAGGGTCAAGCCAGCGCCAGCGCGCTG 873
Db 421 GTGCAGACAGGTTCCTTCCGACCTAATACCCAGGGTCAAGCCAGCGCCAGCGCGCTG 480
QY 874 CTAGTGGCAACATGGGCCCTTCAAGAACCCCGCTGTGTGAGAGCCAGTCAAGGTGCGCTGTG 933
Db 481 CTAGTGGCAACATGGGCCCTTCAAGAACCCCGCTGTGTGAGAGCCAGTCAAGGTGCGCTGTG 540
QY 934 TTTTCTGGGGCCAGCCCACTTTTCTTAAAGGGTGAACAAACAGACATCGGGGTGCG 993
Db 541 TTTTCTGGGGCCAGCCCACTTTTCTTAAAGGGTGAACAAACAGACATCGGGGTGCG 600
QY 994 GGGACTTTCAGATATGCTCTCTAATGGCGGTGCTACCTCGGTGTGGGCGAGTCT 1053
Db 601 GGGACTTTCAGATATGCTCTCTAATGGCGGTGCTACCTCGGTGTGGGCGAGTCT 660
|||||

; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-635-977-1

Query Match 12.2%; Score 272.4; DB 19; Length 1838;
Best Local Similarity 99.6%; Pred. No. 9.8e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 1965
Db 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 2025
Db 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 120

Qy 2026 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 2085
Db 121 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 180

Qy 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGACAGACAGCCAGGACCCAGGAGCCAGAGGAGG 2145
Db 181 TCAGCCACAGAGCTGCAGCAGGAGCAGACAGACAGCCAGGAGCCAGGAGCCAGAGGAGG 240

Qy 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
Db 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

RESULT 10
US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication No. US2004015734A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-9

Query Match 12.2%; Score 272.4; DB 19; Length 1939;
Best Local Similarity 99.6%; Pred. No. 9.9e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 1965
Db 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 2025
Db 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 120

Qy 2026 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 2085
Db 121 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 180

RESULT 11
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US2004017113A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2002-07-09
; PRIOR FILING DATE: 2002-07-09
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match 12.2%; Score 272.4; DB 19; Length 1939;
Best Local Similarity 99.6%; Pred. No. 9.9e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 1965
Db 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 2025
Db 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 120

Qy 2026 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 2085
Db 121 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 180

RESULT 12
US-10-615-659-10
; Sequence 10, Application US/10615659
; Publication No. US2004015734A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10

Query Match 12.2%; Score 272.4; DB 19; Length 1939;
Best Local Similarity 99.6%; Pred. No. 9.9e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 1965
Db 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 2025
Db 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 120

Qy 2026 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 2085
Db 121 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 180

Qy 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGACAGCAAGCCAGGACCCAGAGGAGGAGG 2145
Db 181 TCAGCCACAGAGCTGCAGCAGGAGCAGACAGCAAGCCAGGACCCAGAGGAGGAGG 240

Qy 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
Db 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

RESULT 11
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US2004017113A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2002-07-09
; PRIOR FILING DATE: 2002-07-09
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match 12.2%; Score 272.4; DB 19; Length 1939;
Best Local Similarity 99.6%; Pred. No. 9.9e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 1965
Db 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGAAGTGGGCGGGTTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 2025
Db 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGCAGGCT 120

Qy 2026 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 2085
Db 121 GTGCTGTCTTTTCAAGAAGACTTCGGCGGCAATGGCATCCAGATCTCTCAAGTGGGTGG 180

Qy 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGACAGCAAGCCAGGACCCAGGAGCCAGAGGAGG 2145
Db 181 TCAGCCACAGAGCTGCAGCAGGAGCAGACAGCAAGCCAGGAGCCAGGAGCCAGAGGAGG 240

Qy 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
Db 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

```
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-10

Query Match
Best Local Similarity 6.2%; Score 138.4; DB 19; Length 1859;
Matches 139; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2040 GAAGACTTCCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 2099
Db |||||
Qy 29 GAAGACTTCCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 88
Db |||||
Qy 2100 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2159
Db |||||
Qy 89 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 148
Qy 2160 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2179
Db |||||
Qy 149 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 168
Db |||||

RESULT 13
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10

Query Match
Best Local Similarity 6.2%; Score 138.4; DB 19; Length 1859;
Matches 139; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2040 GAAGACTTCCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 2099
Db |||||
Qy 29 GAAGACTTCCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 88
Db |||||
Qy 2100 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2159
Db |||||
Qy 89 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 148
Qy 2160 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2179
Db |||||
Qy 149 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 168
Db |||||

RESULT 14
US-09-728-446-1073
; Sequence 1073, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
; FILE REFERENCE: LEX-0101-USA
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby
```

```
; CURRENT APPLICATION NUMBER: US/09/728,446
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,270
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1461
; SOFTWARE: RastSeq for Windows Version 4.0
; SEQ ID NO 1073
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Mus musculus
; NAME/KEY: misc feature
; LOCATION: (1)-(452)
; OTHER INFORMATION: n = A, T, C or G
US-09-728-446-1073

Query Match
Best Local Similarity 3.1%; Score 69.2; DB 9; Length 452;
Matches 95; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 132 AGTCTAGTTGGTAAATAATGAGATGCGGTACCTCTGACCATCAAGAGGGACATCA 191
Db |||||
Qy 6 AGTCCCGCATGGTTCGAAATGAGACCCCTACCTCATCTGGACCACTCGGGGGATGTGC 65
Db |||||
Qy 192 TTGACTATCAGAGCTGACCTACGACCATGCTGAACCACTACGCAAGAGACAGCTCT 251
Db |||||
Qy 66 TGGATTGTGCTTCTCTTCCAGGATCAGATGATAAACCATTTATGCCGTCGAGGCTCT 125
Db |||||
Qy 252 TCACCACCAAGGTGAGCC 269
Db |||||
Qy 126 TCACCACCAAGGTGGCC 143
Db |||||

RESULT 15
US-10-450-763-8232
; Sequence 8232, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 8232
; LENGTH: 1042
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SIMILAR
; LOCATION: (346)-(35)
; OTHER INFORMATION: 35% homologous to Streptomyces ansochromogenes SanE, accession
; OTHER INFORMATION: number AF228524, Smith-Waterman Score=100.
US-10-450-763-8232

Query Match
Best Local Similarity 3.0%; Score 67.8; DB 22; Length 1042;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 96 TGAATGCTGGGCATCAATAAGACTTTTCTTCCGAGTCTAGGTGGTAAATAATGAGA 155
Db |||||
Qy 707 TGAATTTGATGACCTAGATGGAACACATGCTCTGTGATGTCCTCCGATGTTCCAGATGAGA 766
Db |||||
Qy 156 TGGCGTACCTCTCTGGACCATCAAGAGGGACATCATGACTATCAGAGCTGACCTACG 215
Db |||||
Qy 767 TCCCTACTTTCATCTGGACCACTCGGGGGATGTGCTCGACTGTGCTTCTCTCCAAAGG 826
Db |||||
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 09:00:19 ; Search time 43 Seconds
(without alignments)
1210.541 Million cell updates/sec

Title: US-10-615-659-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTAEGALRPPPGKGKS #41

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	956.5	33.5	352	2 T12515	hypothetical prote
2	259	9.1	379	2 A45443	tubulin-tyrosine 1
3	252.5	8.8	640	2 E88575	protein ZK1128.6
4	252.5	8.8	680	2 T27699	hypothetical prote
5	209.5	7.3	1198	2 T20262	hypothetical prote
6	209.5	7.3	1203	2 C89217	protein C5SA6.2
7	166	5.8	662	2 T20343	hypothetical prote
8	161	5.6	403	2 T37571	tubulin-tyrosine 1
9	155.5	5.4	883	2 A96805	hypothetical prote
10	144.5	5.1	753	2 S48261	hypothetical prote
11	130	4.6	2472	2 E83594	still frameshift p
12	124	4.3	4660	2 T42737	gp330 protein prec
13	120	4.2	882	2 A70507	probable reductase
14	119.5	4.2	3436	2 S55659	segment protein 6
15	115.5	4.0	1006	2 T41104	probable transcrip
16	114.5	4.0	1336	2 S25716	Ras guanine nucleo
17	112.5	3.9	696	2 T46394	hypothetical prote
18	112.5	3.9	1575	2 S68448	synaptotjanin, 170K
19	111	3.9	930	2 A25923	progesterone recep
20	109	3.8	906	2 A43817	transforming prote
21	108.5	3.8	741	2 I51657	suppressor of yeas
22	108.5	3.8	1048	2 T31425	C-terminal domain-
23	108.5	3.8	1650	2 S53457	dominant autoantig
24	108.5	3.8	2176	2 T39188	probable U5 snRNP-
25	108	3.8	940	2 JG0291	FB19 protein - hum
26	108	3.8	1245	2 G84897	hypothetical prote
27	108	3.8	3149	1 QQB88	BP1F1 protein - hu
28	107.5	3.8	420	2 T46910	hypothetical prote
29	107.5	3.8	896	2 B43817	transforming prote

30 107 3.7 755 2 S32103
31 107 3.7 1234 2 T00363
32 106.5 3.7 538 2 S57459
33 106.5 3.7 2870 2 H96974
34 105.5 3.7 517 2 E89530
35 105.5 3.7 4957 2 T03455
36 105.5 3.7 5262 2 T03454
37 104 3.6 836 2 T32298
38 103.5 3.6 1290 2 T00018
39 103.5 3.6 2774 2 A43359
40 103 3.6 670 2 D86176
41 103 3.6 816 2 T19049
42 102.5 3.6 775 2 S63626
43 102.5 3.6 906 2 JCS963
44 102 3.6 454 2 A84162
45 102 3.6 454 2 T08280

ALIGNMENTS

RESULT 1

T12515

hypothetical protein DKFZp434B103.1 - human

C;Species: Homo sapiens (man)

C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004

C;Accession: T12515

R;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A;Reference number: Z17524

A;Accession: T12515

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-352 <WAM>

A;Cross-references: UNIPROT:Q9Y4R7; EMBL:AL096725

A;Experimental source: adult testis; clone DKFZp434B103

C;Genetics:

A;Note: DKFZp434B103.1

Query Match 33.5%; Score 956.5; DB 2; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.2e-64;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSDADVEDITEAFWEDLTQYYSLVHGDAFISNRVYFSOCALLANRITVSNQTDI 133
Db 2 DIDKDLAEPLYTPEGWSLFLQRYQQVHGEAELRLDTQVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNIIWKPAKSRGRDIVCMRDVVEEILEAAADHPLSRDNKWKVVKYIETPLLICDT 193
Db 62 EGDRIWIIVKPKSKRGIMCMHLEMLKLVNGPNVVKDGKWKVVKYIERPLLIFFT 121
QY 194 KFDIQQWFLVTQNNPLTIWFKESYLRFTQRPFLSKDKLSAHLCCNNAVKYLRNDVGRS 253
Db 122 KFDLRQWFLVTQNNPLTIWFKESYLRFTQRPFLSKDKLSAHLCCNNAVKYLRNDVGRS 181
QY 254 FLIPAHNMVTSRFQYLRQGRGAVNGSVIYPSMKKATAHAKMVAQDHVPEKNSFELY 313
Db 182 PLLPPNNMSSORFQAHLEMGAPNAWSTIIVFGMKDAVIHALQTSQDTVQCRKASFELY 241
QY 314 GADFVLGRDFRPLWLEINSPTMHPSTPTAQICAOVQEDTIKAVV----DRSCDIGNFE 369
Db 242 GADFVGEDEFQPLWLEINASPMPASTATATRLCAGVQADTLRVVIDRMLDRNCDTGAFE 301
QY 370 LLWRQPVVEPPPPFSGSDCLCAGV-SVRRAR 398
Db 302 LIYKQPVTTSPASTPRPSCLLPMSYDTRAR 331

RESULT 2

A45443

tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A45443
R;Brsfeld, K.; Wehland, J.; Plessmann, U.; Dödemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A;Title: Characterization of the tubulin-tyrosine ligase.
A;Reference number: A45443; MUID:93147125; PMID:8093886
C;Accession: A45443
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-379 <ERS>
A;Cross-references: UNIPROT:P38160; GB:X68453; GB:S54050; NID:q2199; PIDN:CAA48494.1; PID:3881506; GSPDB:GN00021; CESP:ZK1
A;Note: Sequence extracted from NCBI backbone (NCBIN:123858, NCBI:P:123859)
C;Keywords: ligase

Query Match 9.1%; Score 259; DB 2; Length 379;
Best Local Similarity 26.8%; Pred. No. 1e-11;
Matches 98; Conservative 57; Mismatches 128; Indels 82; Gaps 20;

QY 52 PG--QLVDI---ACKVC-QAYLQLEHEDIDTSADAVEDLTAEWEDLTQYYSL-----100
DB 56 PGLMQLVNYRGADKLCRKASLVKL----IKTSPELAEST---WFPESVVIYPTNLKTP 108
QY 101 -----VHGDAFISNRYFSQCOALLNRITTSVNPQTDIDGLRNIWIIPAKSR 149
DB 109 VAPAQNGIHPPIHSSR--TDREFF-----LTSYNKKKE-DGEGNVWIAKSSAGAK 156
QY 150 GRDIVCMRVETLELAADHPLSRDNK--WVQKYIETPLLI--CDTKFDIROWFLVLT 204
DB 157 GEGILSSSEATELDFI-----DNQGVHVIQKTLERPLLEPCHRKFDIRSWLVD 208
QY 205 DWNPLTIWFKESYLRFSQFSLDKL-DSAIHLCNNAVQKYLKNDVGRSPLLPAAHNMWT 263
DB 209 --HOYNIYLYREGVLRTASEPHYTDNFQDKTCHLTWHCIQKEYSKNYGK--VEEGNEWF 263
QY 264 STRFQYLRQGRGAVGWSVTPSMKKAIAHAMKVAQDHVEPR-----KNSFELYGAD 316
DB 264 FEEFNQYLT-----SALNTLESSILLOIKHIRSCLLSVEPAISTRHLFPYQSFQFGFD 318
QY 317 FVLGRDFRPLWLEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVDRSCDIGNFELLWRQPV 376
DB 319 FWVDEDLKWLIEVNGAPAC--AQKLYAELC-----QGIVDIAIASVFPFPDAB-----QQQ 368
QY 377 VEPPPP 381
DB 369 QQPPP 373

RESULT 3
E88575
protein ZK1128.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: E88575
R;Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88575
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-640 <STO>
A;Cross-references: UNIPROT:Q09647; GB:chr_III; PIDN:CAA87425.1; PID:g3881506; GSPDB:GN00021; CESP:ZK1
A;Gene: ZK1128.6
A;Map position: 3

Query Match 8.8%; Score 252.5; DB 2; Length 640;
Best Local Similarity 25.6%; Pred. No. 6.8e-11;
Matches 81; Conservative 56; Mismatches 118; Indels 61; Gaps 12;

QY 132 DIDGLRNIWIIPAKSRGRDIVCMRVETLELAADHPLSRDNKVVQKYIETPLLIC 191
DB 323 ETDSRHHV-IVKPPASARGTGISVTRKPKDPPTATL-----VAQHVIERPLTN 371
QY 192 DTKFDIROWFLVTDWNPLTIWFKESYLRFSQFSLDK---LDSAIHLCNNAVQKYLK 248
DB 372 RAKFDRLIYAYVPTPEPLRVYIYDQGLVRFSVSVSHSVSTISNKMHLTNTYSINKLA 431
QY 249 D-VGRSPL--LPAHNMWTSTR-----QEYLQROGRGAVGWSVTPSMKKAIAHAMKVA 299
DB 432 DGVANKPVPKWTLLHLW--EHPDEMGVDREKIQRE-----IEEVIKAFISTEKPI 480
QY 300 QDH-----VEPRKNSFELYGADFLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTI 355
DB 481 REHMSRFLQEFCYELFGIDILDEYKPMLEVNISPLSHSGTPLDVSVKAPLAKDVL 540
QY 356 KVAVDRSCDIGNFELLWRQPVVEPPPF---SSSDLCVAGSVRRARRQVLP-----VCNLK 408
DB 541 NILA-----GVYVPPSFDKLSADYSTPRNGRKNRQLEIKEASWVAAYK 584
QY 409 ASASLLDAQPLKARGP 424
DB 585 DQLGVIDNRIFKRLTP 600

DB 283 ETDSRHHV-IVKPPASARGTGISVTRKPKDPPTATL-----VAQHVIERPLTN 331
QY 192 DTKFDIROWFLVTDWNPLTIWFKESYLRFSQFSLDK---LDSAIHLCNNAVQKYLK 248
DB 332 RAKFDRLIYAYVPTPEPLRVYIYDQGLVRFSVSVSHSVSTISNKMHLTNTYSINKLA 391
QY 249 D-VGRSPL--LPAHNMWTSTR-----QEYLQROGRGAVGWSVTPSMKKAIAHAMKVA 299
DB 392 DGVANKPVPKWTLLHLW--EHPDEMGVDREKIQRE-----IEEVIKAFISTEKPI 440
QY 300 QDH-----VEPRKNSFELYGADFLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTI 355
DB 441 REHMSRFLQEFCYELFGIDILDEYKPMLEVNISPLSHSGTPLDVSVKAPLAKDVL 500
QY 356 KVAVDRSCDIGNFELLWRQPVVEPPPF---SSSDLCVAGSVRRARRQVLP-----VCNLK 408
DB 501 NILA-----GVYVPPSFDKLSADYSTPRNGRKNRQLEIKEASWVAAYK 544
QY 409 ASASLLDAQPLKARGP 424
DB 545 DQLGVIDNRIFKRLTP 560

RESULT 4
T27699
hypothetical protein ZK1128.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27699
R;Berks, M.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z20407
A;Accession: T27699
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-680 <MIL>
A;Cross-references: UNIPROT:Q09647; EMBL:247357; PIDN:CAA87425.2; GSPDB:GN00021; CESP:ZK1
A;Experimental source: clone ZK1128
C;Genetics:
A;Gene: CESP:ZK1128.6
A;Map position: 3
A;Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2

Query Match 8.8%; Score 252.5; DB 2; Length 680;
Best Local Similarity 25.6%; Pred. No. 7.4e-11;
Matches 81; Conservative 56; Mismatches 118; Indels 61; Gaps 12;

QY 132 DIDGLRNIWIIPAKSRGRDIVCMRVETLELAADHPLSRDNKVVQKYIETPLLIC 191
DB 323 ETDSRHHV-IVKPPASARGTGISVTRKPKDPPTATL-----VAQHVIERPLTN 371
QY 192 DTKFDIROWFLVTDWNPLTIWFKESYLRFSQFSLDK---LDSAIHLCNNAVQKYLK 248
DB 372 RAKFDRLIYAYVPTPEPLRVYIYDQGLVRFSVSVSHSVSTISNKMHLTNTYSINKLA 431
QY 249 D-VGRSPL--LPAHNMWTSTR-----QEYLQROGRGAVGWSVTPSMKKAIAHAMKVA 299
DB 432 DGVANKPVPKWTLLHLW--EHPDEMGVDREKIQRE-----IEEVIKAFISTEKPI 480
QY 300 QDH-----VEPRKNSFELYGADFLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTI 355
DB 481 REHMSRFLQEFCYELFGIDILDEYKPMLEVNISPLSHSGTPLDVSVKAPLAKDVL 540
QY 356 KVAVDRSCDIGNFELLWRQPVVEPPPF---SSSDLCVAGSVRRARRQVLP-----VCNLK 408
DB 541 NILA-----GVYVPPSFDKLSADYSTPRNGRKNRQLEIKEASWVAAYK 584
QY 409 ASASLLDAQPLKARGP 424
DB 585 DQLGVIDNRIFKRLTP 600

```
RESULT 5
T20262
hypothetical protein C55A6.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T20262
R;Kershaw, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19243
A;Accession: T20262
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1198 <WIL>
A;Cross-references: UNIPROT:O17720; EMBL:Z81051; PIDN: CAB02862.2; GSPDB: GN00023; CESP: C5
A;Experimental source: clone C55A6
C;Genetics:
A;Gene: CESP: C55A6.2
A;Map position: 5
A;Introns: 32/3; 68/1; 107/3; 201/3; 273/1; 364/3; 407/3; 460/3; 476/2; 534/2; 677/3; 73

Query Match 7.3%; Score 209.5; DB 2; Length 1198;
Best Local Similarity 25.5%; Pred. No. 2.9e-07;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;

Qy 140 WIIKPAKSRGRDIVCMRVVEILEAAADHPLSRDNKWWVQKVIETPLICTKFDIRQ 199
Db 715 FIVKPTNSRQKGIFANSWADI-----PARGPL-----LVSRYLKDPYLVNNHKFDLRI 764

Qy 200 WFLVTDNPLTIWFYKESYLRFSTORF--SLDKLDS--AIHLCNNAVOK----YLKNDVG 251
Db 765 YVAVTSFYPLVAVYSGLARLASRPYDTSASSADSNEYVHLTNYSINKNSTSFVRNESM 824

Qy 252 RSPLLPAHNMWTTTRFOEYLQROGRGAVGWSVIYPSMKKAIAMKVAQDHV--EPRKN- 308
Db 825 SSEDLL--GHKWTLLGALLRYVENEKDA---KLLMLRIEDLIVKSLLSIGNSVATASRTNL 879

Qy 309 -----SPELGADFLVGRDRFPMLEINSPTMHPSTPVTQAQLCAQVQEDTIKVA 358
Db 880 RFACITNPELFGFDVLVDQALKPMLLEVLNPSLACDAPLDSLKLTRLIADLLNLA 934

RESULT 6
C89217
protein C55A6.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: C89217
R;Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID: 99069613; PMID: 9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: C89217
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1203 <STO>
A;Cross-references: UNIPROT: O17720; GB: chr_V; PIDN: CAB02862.1; PID: g3875273; GSPDB: GN000
C;Genetics:
A;Gene: C55A6.2
A;Map position: 5

Query Match 7.3%; Score 209.5; DB 2; Length 1203;
Best Local Similarity 25.5%; Pred. No. 2.9e-07;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;

Qy 140 WIIKPAKSRGRDIVCMRVVEILEAAADHPLSRDNKWWVQKVIETPLICTKFDIRQ 199
Db 720 FIVKPTNSRQKGIFANSWADI-----PARGPL-----LVSRYLKDPYLVNNHKFDLRI 769

Qy 200 WFLVTDNPLTIWFYKESYLRFSTORF--SLDKLDS--AIHLCNNAVOK----YLKNDVG 251
Db 770 YVAVTSFYPLVAVYSGLARLASRPYDTSASSADSNEYVHLTNYSINKNSTSFVRNESM 829
```

```
Qy 252 RSPLLPAHNMWTTTRFOEYLQROGRGAVGWSVIYPSMKKAIAMKVAQDHV--EPRKN- 308
Db 830 SSEDLL--GHKWTLLGALLRYVENEKDA---KLLMLRIEDLIVKSLLSIGNSVATASRTNL 884

Qy 309 -----SPELGADFLVGRDRFPMLEINSPTMHPSTPVTQAQLCAQVQEDTIKVA 358
Db 885 RFACITNPELFGFDVLVDQALKPMLLEVLNPSLACDAPLDSLKLTRLIADLLNLA 939
```

```
RESULT 7
T20343
hypothetical protein D2013.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T20343; T22085
R;Mortimore, B.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z19259
A;Accession: T20343
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-662 <WIL>
A;Cross-references: UNIPROT: Q09512; EMBL: Z47808; PIDN: CAAB7778.1; GSPDB: GN00020; CESP: D2013.9
A;Experimental source: clone D2013
R;Matthews, P.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z19510
A;Accession: T22085
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-662 <WIL2>
A;Cross-references: EMBL: Z47809; PIDN: CAAB7783.1; GSPDB: GN00020; CESP: D2013.9
A;Experimental source: clone F42A8
C;Genetics:
A;Gene: CESP: D2013.9
A;Map position: 2
A;Introns: 78/2; 114/3; 330/3; 471/1; 521/2; 587/2

Query Match 5.8%; Score 166; DB 2; Length 662;
Best Local Similarity 24.3%; Pred. No. 0.00022;
Matches 72; Conservative 45; Mismatches 115; Indels 64; Gaps 14;
```

```
Qy 78 SADAVEDLTAEWEDLTQQYISL-VHGDAFISNRYFSQCALLNRITSVPOTDIDGL 136
Db 396 AACAMRDPKNDYQLT---YNLNTQLPEFVAFQN-----RELNGQ 434

Qy 137 RNWIIKPAKSRGRDIVCMRVVEILEAAADHPLSRDNKWWVQKVIETPLLI----- 190
Db 435 HNWIVKPMNLARGMDMTVTEDLNQIRMI-----ETGPKIVCEYIPRPLPPRPDNG 487

Qy 191 CDTKFDIRQWFLVTDNPLTIWFYKESYLRFSTQFSLDKL-DSAIHLCNNAVQKYLKND 249
Db 488 NKVKFDLRIYVFLNGAPVATYVYVNFWRFAINFSLNSFEDVETHF---TVENYL--D 542

Qy 250 VGRSPLLPAAHNMWTTTRFOEYLQROGRGAVGWSV---IYPSMKKAI-AMKVAQDHVEP 305
Db 543 KEKILQMKCN-----FIETIEKAPRIQWSEVQKDINLTIRKAIKAEAAKEAPRGVAP 596

Qy 306 RKNSELYGADFLVGRD-----FRPWLIEINSSPTMHPSTPVTQAQLCAQVQEDTI 355
Db 597 NVQSRAMYGVIMLQHGDNNDVIKSTLLEINFMPD-----TTRACQYYPDFADTV 645
```

```
RESULT 8
T37571
tubulin-tyrosine ligase related protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T37571
R;Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z21726
```

A:Accession: T37571
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-403 <BAD>
A:Cross-references: UNIPROT:Q10438; EMBL:Z70721; PIDN:CAA94694.1; GSPDB:GN00066; SPDB:SE
A:Experimental source: strain 972h-; cosmid c12B10
C:Genetics:
A:Gene: SPDB:SPAC12B10.04
A:Map position: 1
A:Introns: 320/3; 348/3

Query Match 5.6%; Score 161; DB 2; Length 403;
Best Local Similarity 20.6%; Pred. No. 0.00025;
Matches 72; Conservative 65; Mismatches 127; Indels 86; Gaps 14;

QY 71 EHEIDTSADAV-----EDITEAWEDLTQQYVSLVHGDAFTNS----- 110
DB 49 QYEDID--FDEVKNPKTKLCCSVIRKALIRKEYLWRVITVYLAHPDSILSKSVPEAY 106

QY 111 ---RNYFSCQALLNRITSVNPQTIDGLRNI-----WIIPKAASGRGRDIVCMRVERI 162
DB 107 SLELDVYAEFLDSDLSMEAYELRQLEENATKNISEKQWYILKPSMCDRAQGIRLFSFIEEL 166

QY 163 LEL-----AAADHPLSRDNK-----WVVKIETPLIICDTKFDI 197
DB 167 QAIFDSFDDSESESEAGLEEKGDIITAFNNKIVISQIRNFLVKYISKPLLLDHRKPHI 226

QY 198 RQWFLVTDWNPITWFKESYLFQSFSLDKLDSAI---HLNNAVQYKLNKDVGRSP 254
DB 227 RAYVLAT--GALSYVLFNEMCLLARDKYKTPDPDLFLSHLSNTCLQ---GDNVEQSS 281

QY 255 LLPAHNMWTSRFOEYLQQRGAVGWSVITYPSMKKAIAHAMKVAQD-----HVEPRKNSF 310
DB 282 I---RDFWNTS-----IENK-----DDIPKSLNIIGDVFEAAATTQGIHFQPLENCF 326

QY 311 ELYGADFLVGRDPRFWLIEINSPTWHPSTPTVAQLCAQVQEDTIKAVD 360
DB 327 EIFGVDFLVDCESQVYLLEVS-----YPDFKQTGKNLSNIENLFSAVVE 372

RESULT 9
A96805
Hypothetical protein TSM16.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96805
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, W.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96805
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-883 <STO>
A:Cross-references: UNIPROT:Q9CAQ1; GB:AE0051173; NID:g3382502; PIDN:AAF07788.1; GSPDB:GN
C:Genetics:
A:Gene: TSM16.14
A:Map position: 1

Query Match 5.4%; Score 155.5; DB 2; Length 883;
Best Local Similarity 25.3%; Pred. No. 0.0021;
Matches 55; Conservative 36; Mismatches 93; Indels 33; Gaps 7;

QY 134 DGLRNWIIKPAASGRDIVCMRVERIELEAAADHPLSRDNKVVVQKIETPLIICDT 193
DB 641 DQNNLWILKPNWNTIDTISITDNLSAIR-----MMETGPKTCQKIIEHPALFKGN 693

QY 194 KFDIROWFLVTDWNPITWFKESYLFQSFSLDKLDSAIHLNNAVQYKLNKDVGRS 253
DB 694 KFDLRVVLVRSIDPLEIYLIEIFWVRLSNPNPSLEKHSFFEYETHFTVMNY-----GRK 748

QY 254 PLLPAHNMWTSRFOEYLQQRG-----RGAVGWSV-----IYESMK---KAIAHAMK 297
DB 749 ----LNHKPTABFVRFEQEHNDYAFHFVNNTIQLSIVKWMMDIHEKVKQVIRAVFEAAA 804

QY 298 VAQDHVEPRKNSFELYGADFLVGRDPRFWLIEINSSP 334
DB 805 LAHPMQSPK--SRAMYGVDVMDLSDSFEPKILEVTYCP 840

RESULT 10
S48261
Hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR0821
C:Species: Saccharomyces cerevisiae
C>Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004
C:Accession: S48261; S45962; S41800; S44676
R:Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A:Reference number: S48255; MUID:95208357; PMID:7900426
A:Accession: S48261
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-753 <MAN>
A:Cross-references: UNIPROT:P38254; EMBL:X78993; NID:g476045; PIDN:CAA55599.1; PID:g4760;
R:Feldmann, H.; Mannhaupt, G.; Schwarziöse, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45962
A:Molecule type: DNA
A:Residues: 1-753 <FE2>
A:Cross-references: EMBL:Z35963; NID:g336366; PIDN:CAA85047.1; PID:g536367; MIPS:YBR094w
R:Dekker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
submitted to the EMBL Data Library, December 1992
A:Reference number: S31313
A:Accession: S41800
A:Molecule type: DNA
A:Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
A:Cross-references: EMBL:X69881; NID:g4086; PIDN:CAA49508.1; PID:g4088
C:Genetics:
A:Cross-references: SGB:S0000298
A:Map position: 2R

Query Match 5.1%; Score 144.5; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 0.011;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

QY 140 WIIPKAASGRDIVCMRVERI-----LELAADHP----- 171
DB 476 WTVKPSMSDKGGIRVFKTIEDQLAIFDSDSEAESGNDSDADDVNGEFMDNKNVN 535

QY 172 LSRDNKVVVQKIETPLIIC---DTKFDIROWFLVTDWNPITWFKESYLFQSFQRF-- 226
DB 536 ISQLRHFIIQEVLTPNPLLASMDNKRKFIKCY--VVCGRDLQVVFVYDRMLALPAKPFVP 593

QY 227 -----SLDKLDSAIHLNNAVQYKLNKDVGRSPLLPAHNMWTSRFO--EYLQQRG 277
DB 594 LDPYAVSVTDLKDLECHLTNTCLQS--KKDKDSSVL-----EFDISIEIPNERS 642

QY 278 AVWGSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFLVGRDPRFWLIEINSSPTMH 337
DB 643 NT-KEQIHSITNDVFLAAVNVNRLNFQPLNAPFETYGVDFLIDSNYEYVKKLEINAFPDFK 701

QY 338 PSTPVTVAQLCAQVQEDTIKAVV 359
DB 702 QTGDKLKNLIDELFDDTVKVCV 723

RESULT 11

E83594
still frameshift probable component of chemotactic signal transduction system PA0413 (im
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: E83594
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Ba
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83594
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2472 <STO>
A;Cross-references: UNIPROT:Q91696; GB:AE004478; GB:AE004091; NID:99946261; PIDN:AAG0380
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA0413

Query Match 4.6%; Score 130; DB 2; Length 2472;
Best Local Similarity 19.7%; Pred. No. 0.76;
Matches 122; Conservative 81; Mismatches 203; Indels 214; Gaps 29;
QY 71 EHEDITD-----SADAVEDLTAEWEDLTQOYSLVHGDAFISNSRNYFSCQALLN 122
DB DHEVTDAERARLAGPRDAMRSVVGALCEELVRIKDSL---DLFVRSDRGHPSELALLA 354
QY 123 RITSVNPQTIDG-----LRNIWIKPAKSRGR--DIVCMRVETEELEAA----- 167
DB 355 PLQAIADTLAVLFGQPRKVLQDLQVHALAGREPSDAILMDVAGALLYVEATLAGM 414
QY 168 ---ADHPLSRDNKVVQKYIET-PLLICDTKFDIRQ-----WFLVTDNMPLTIWFYKE- 216
DB 415 AGPGDERNSSESLPTTDVAQIHLQVKEARNGLEQAKDAIEFIASQWNHEHLARVPEL 474
QY 217 -SVLRSTORFSLDKLSAHLGNNAVOKYKNDVGRSPLLPANHMTS-----TRF 267
DB 475 LTQVRGLAMIPLERAATLTETCNRYIQEQL---LARK---AVPDWQSLDTLADAITSV 527
QY 268 QEYLQQRGCVGWSVTPYSMKKAIAHAMKVAODHVEPRKNSFELYGADPVLGRDFRPL 327
DB 528 EYLERL-----SEDHASQSDLI---LDVAEDSLAN-----LGYTLKP-- 562
QY 328 IEINS-----SPTMHPSTPVTQAQLCAQ-----VOEDTIKVA 358
DB 563 ---NSSAPAEPLGSLGPAIESPAEPEPAVVEVAETAEPADTAPAEAREADAPOLA 619
QY 359 VDRSCDIG-----NPELLWRQPV---VEPPPP-----F 382
DB 620 SDNNTLGEVAPDAGEPFLDLDLPLDSDSAEVPFALPEVVEESGQPSQTSAPARSLDDF 679
QY 383 SGSDLCVAGSVRRARRQVLPVCNKLKASALLDAQPLKARGPSAMPD----- 429
DB 680 SLDEIDLGLD-----LPA-----DAAP--ASGPALADWSLPEQWGLGDDL 719
QY 430 --PAQGPSS-----PALQDRLGLKEEKGKPLALLAPLRAAESG---GAAQP-----T 472
DB 720 AQAQAQAGETLDSLRLPALSFDPAPLESLE--PLPALEFPDGSAAEQELVLDALDPLDVA 777
QY 473 RTAAKGV-----ELPACPCRHVDSQAPNTGVVPAQPAKSWDPNQLNAH 516
DB 778 LPPEAGEVSAWEGSSLEELDLSLDLPEVQLPEAEAEAPPAAEALASEAPALSVAEMAA 837
QY 517 PLEPVLRLGLTAEGALRPPPP 536
DB 838 PVQPIPPPAQNVPSLLPPP 857

RESULT 12

T42737

gp330 protein precursor - rat

N;Alternate names: megalin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42737
R;Saito, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A;Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of th
A;Reference number: A58173; MUID:95024033; PMID:7937880
A;Accession: T42737
A;Status: preliminary; translated from GB/EMBL/DBU
A;Molecule type: mRNA
A;Residues: 1-4660 <SAI>
A;Cross-references: UNIPROT:P98158; EMBL:L34049; NID:9561852; PID:9561853; PIDN:AAA51369.
A;Experimental source: strain Sprague-Dawley; kidney
C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
F;1-25/Domains: signal sequence #status predicted <SIG>
F;26-4660/Product: gp330 protein #status predicted <MAT>
Query Match 4.3%; Score 124; DB 2; Length 4660;
Best Local Similarity 19.7%; Pred. No. 5.4;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;
QY 66 YLGQLEH-EDIDTSADAVEDLTAEWEDLTQOYYS-LVHGDAFISNSRNYFSCQALLNR 123
DB 4079 YLEEEHEIQTIDYDWD-----PEHIGLSVVYTVLAQGSQFGAIKRAYIPNFESGSN- 4130
QY 124 ITSVPNTQDID-GLRNIWIKP---AAKSRGRDIVCMRVETEELEAAADHPLSRDNKVV 179
DB 4131 ---NPIREVDLGLK--YLMQPDGLAVDVGRHIYMSDAKSQRIEVATLD---GRYRKWL 4181
QY 180 VKQYIETPLLICDTKFDIRQWFLVTDNMPLTIWFYKESYLRFSTQFSLDKLSAHLN 239
DB 4182 ITTQLDPPAAIA-----VNP-----KGLMFWTDQKQKRIESA----- 4215
QY 240 NAVQYKLNKDVGRSPLLPANHMTSTRFQBYLQQRGAVGWSVTPYSMKKAIAHAMK-- 297
DB 4216 -----WMNGEHRSLVSENGLWPNGLSIDYLNDD-----RVYMSDSKEDVIEALKYD 4262
QY 298 -----VAQDHVEP-----RNSFELYGADPVLGRDFRPLWIEI 330
DB 4263 GTDRRLIINEAMKPFSLDIFEDKLYWVAKEGKSWVRQKNGKENVL--VVPMLTQV 4320
QY 331 N-----SPTMHPSTPVTQAQL-----CAQVEDTIKVAVDNRSCDIGNFELLWRQ 374
DB 4321 RIFHLRYNQSVNSPCKVCSHLCLLRPGYSCACPGSDFTVGTSTVQCDAAS-EL----- 4375
QY 375 PVVEPPPP-----VCNKLKASALLDAQPLKARGPSAMPDPAQG----- 433
DB 4376 PVTMPPPPCRMHGNCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 4429
QY 404 -----VCNKLKASALLDAQPLKARGPSAMPDPAQG----- 433
DB 4430 LTFVIVIIVGALVLVGLFHYRKTGSLTLTP-LKPLSLSLAKPSNGNGVTFRSGADVNM 4488
QY 434 -----PSPALQRLDGLKE---EKG-LPLALLAPLRAEAESGGAAQPTRTKAAGKVE 481
DB 4489 DTGVSPFGPETIIDRSMANNEHFVMEVGQPVIFENPMYAAKDN-----TSKVA 4537
QY 482 LPACPCRHVDSQAPNTGVVPAQPAKSWDPNQLNAHPLVRLGLKTAEGALRPPPGKGS 541
DB 4538 LAV-----QGPSTGAQVTPENV--ENQNYGRPIDP-----SEIYPEKPPASPGA 4580
RESULT 13
A70507
probable reductase (EC 1.3.99.-) iron-sulfur protein - Mycobacterium tuberculosis (strain
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: A70507
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A;Authors: Sqaes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70507
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-882 <COL>
A;Cross-references: UNIPROT:O32368; GB:Z97991; GB:AL123456; NID:g3261837; PIDN:CAB10793
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0338c
C;Keywords: oxidoreductase

Query Match 4.2%; Score 120; DB 2; Length 882;
Best Local Similarity 27.2%; Pred. No. 0.95;
Matches 43; Conservative 11; Mismatches 52; Indels 52; Gaps 5;

QY 416 AQPLKARGPSAMP-----PAQGSPPALORDLGLKEEKLPLALLAPLRGNAES 455
 :
DB 736 AKQAQERAPKAAPKAAAPVTFVEAPAEPQAPA-----PAAAPAAPVKGLGMA 782
 :
QY 466 GGAAOPTRTKAAAGVELPACPGRHVD-----QAPNTGVPPVAQPAKSW 508
 :
DB 783 AGAKRPQAKAAATTPAAPAPAPAPVPKGLGTAGACRRCAKTKTTPPAPGLAEPAAPQPEA 842
 :
QY 509 DPNQLNAHPLEP-----VLRLKTAEGALRPP 535
 :
DB 843 KPQEPAPAPPKPOTDGDPAAPAPAPVPKGLGTARGA-RPP 879
 :

RESULT 14
S55659
segment protein 64 - equine herpesvirus 2
C;Species: equine herpesvirus 2
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C;Accession: S55659
R;RefSeq: E.A.R.: Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A;Title: The DNA sequence of equine herpesvirus 2.
A;Reference number: S55594; MUID:95302501; PMID:7783207
A;Accession: S55659
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3436 <TEL>
A;Cross-references: UNIPROT:Q66666; GB:U20824; NID:g6951172; PIDN:AAC13852.1; PID:g695237
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match	4.2%	Score 119.5;	DB 2;	Length 3436;
Best Local Similarity	18.6%;	Pred. No. 7.5;		
Matches 134;	Conservative 94;	Mismatches 261;	Indels 233;	Gaps 29;
Qy	9	VVSHQSCSRSSRSKPRDREAGS----	SDLSSRQDAENAEAKLGLPQLQVLDIACKVC	63
Db	1919	VATGSLUPFLRHSPAPCAEDTGAALFALSEAVKLEKSGKRIWRWGEVGGKLSDAHSLVA	1978	
Qy	64	QAYLGLQLEHEDI-----DTSADAVEDLTAEAWEDLTQQYYSLVHGDAPFISNRNYF	114	
Db	1979	RSAAKGAQKRFALIOQLRKDVAAQREIME-DWK-----AFVT-----	2018	
Qy	115	SQCCALNLRITSNPQTDIDGLRNWIIKPAKSRGRDIVCMDBRVEIILELAADHPLSR	174	
Db	2019	---EAPINSMEDVND-----LRAAPNEAEFEPAKLEBRKVRMREAEAKLEK	2064	
Qy	175	DNKVVVQKYYTETPLLIC-----DTKFIQIRQFLVTT--DWNPLTITWFVYESYLRFSTORFSL	228	
Db	2065	EAEEMLTQAVKRGRLQCGWRIQNAFDNMAFGGITGEDWAAVAAAEFQREGTSLSTLPQGQ	2124	
Qy	229	DKLDSAIH-----LCNNNAVOKYLKNDVGRSP-----LLP--AH-NMWTSTRFQEYLQROG	275	
Db	2125	SKLTDKVEAEVEALLNKNVSMPLNGPAPFKPPAFDWLTTPYRAHMNFYFKLSFPPLPKLNROA	2184	
Qy	276	RG-----AVWGSGVIYPS-----MKCAIAHAMKV-----AQDHVEPRKNSFELYGADF	317	

Db	2185	EAVEAKMSQTEQAIEGADVTEAVAGTPLEAPVARALRLLEAARDEAAGLKQIDEGE	RAY	224
Qy	318	VLG-----RDFRPWLIIE-----	329	
Db	2245	VQVERAGEGEPKAEIPKKLLTYEQTLSLANLPEDFQKNVLQNETLMLNQLREYL	2304	
Qy	330	-----INSPTWHPST--PVTALCAQOEDTIKVAV-----DRSCDIGNFELLWR--	373	
Db	2305	GRVTENINSLESRAKTSRGEANRLAAVIEENIPQANVSISSRRLDKSDPVGFLEGI	2364	
Qy	374	QPVVEPPPPSGSDLCVAGVSVRRARQVLPVC--NLKASASLLDAQPLKARGPSAM	431	
Db	2365	KQIVESDPYSATRESL--VWLHRTFKALLPLCPASLKRMMELIGEEITLREKG	2416	
Qy	432	QGPPSPALQ-----RDLGLKE--	447	
Db	2417	RGAPRVCVRAHETDDGVLTETLEARVTCGKAAVEGVKKRDYARNKMWEDLAL	2476	
Qy	448	-EKGL-PL-----ALLAPLRGAESGGAAQPTRTTKAAGKVLPACPCRH-	489	
Db	2477	VEKRLGFLVKSREALDAPDLAVLQEEAALLLEAKTGGLDK--SAPETHERVLELQ	2534	
Qy	490	-----VDSQAPNTGVVPVAPQAPKSWDPNQLNAHPLEPVLRLKTAEGALRP	539	
Db	2535	RFKDLFLKHYLDSQRPVFEEAPLSRALYWSNQSQERESRENSGEGEGEGERR	2594	
Qy	540	GS	541	
Db	2595	GS	2596	
RESULT 15				
T41104				
probable transcription factor subunit, TPR domain - fission yeast (Schizosac				
C;Species: Schizosaccharomyces pombe				
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004				
C;Accession: T41104				
R;Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.				
submitted to the EMBL Data Library, September 1998				
A;Reference number: Z21964				
A;Accession: T41104				
A;Status: preliminary; translated from GB/EMBL/DBJ				
A;Molecule type: DNA				
A;Residues: 1-1006 <P>R>				
A;Cross-references: UNIPROT:O74458; EMBL:AL031535; PIDN:CAA20753.1; GSPDB:GN000				
A;Experimental source: strain 972h-; cosmid c16C4				
C;Genetics:				
A;Gene: SPDB:SPCC16C4.14c				
A;Map position: 3				

Query Match	4.0%; Score 115.5; DB 2; Length 1006;
Best Local Similarity	21.0%; Pred. No. 2.5;
Matches	92; Conservative 58; Mismatches 137; Indels 151; Gaps 24;
Qy	28 EAAGSDLSRQD-----AENAEAKRLGPG--QLVDIAACKVCOA---YLGQLEHEDI 75
Db	3 QNGNSVDSNMNETQNTDTDNFDEAQDLNGYISEIVDEARNVSEVDKFLG----- 55
Qy	76 DTSADAVEDLTEABWEDLTQOYYSLVHGDAFINSRNYFSQCOALLNRITSVNPQTDI-- 133
Db	56 DTSALQAEGL---WSDEESDYEG-----SDDESNSFSK-----TASTEDDIAN 95
Qy	134 -----DGLRNIWIIKPAAKSRGR-----DIVCMDRVEEILEAAADHPLSRDNKW- 178
Db	96 EWEENLKA VAGFRK--VRKGHGKGRVSRA DMLPSVEVQOMLSL--ANHFLAQEGNFD 150
Qy	179 VWQYIETPLLICDTKFDIPQWFLVTDWNP LTTW-FYKESYLFSTQRFSLDKL-----DS 233
Db	151 EAQLKAEIIVRI-----DNNVIAAKWMLGECHRGNGRNVNIEKLIAMWA 196
Qy	234 AIHLCNNNAVOKYLKNDVGRSPLLPAHNMW--TSTRFOYLROQGRGAVWMSGVIY----- 285
Db	197 AAHL-----KP--KHLELWPTCAKLSSELS-----FWDQADYCNRAVS 233

```
Qy 286 -----PSMKKAI AHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPWLIEINSSPTWHPS 339
Db 234 AKPPKSELKKYIWNRSVLNKEHGLKAA---EGFKFL-----QS 272
Qy 340 TPVTAQLCAQVQEDTIKAVDRSCDIGNFELLWRQPVVEPPFSGSDLCVAGVSVRRARR 399
Db 273 SPYNASILKNLAETIYIKIHAPREI-LKQFEIAWKYFYQYPAPPIGNDIF----- 320
Qy 400 QVLPVCNLKASASLDDAQ 417
Db 321 -DLPTNLNLYAEILLLDHQ 337
```

Search completed: September 17, 2005, 09:11:22
Job time : 46 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 08:38:12 ; Search time 43 Seconds
(without alignments)
939.189 Million cell updates/sec

Title: US-10-615-659-2
Perfect score: 2854
Sequence: 1 MASSILKVVYSHQSCRSSR.....LRGLKTAEGALRPPGGKGS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	212	7.4	561	4	US-09-270-767-46703
2	154.5	5.4	260	4	US-09-270-767-62411
3	154.5	5.4	507	4	US-09-270-767-46787
4	151	5.3	259	4	US-09-248-796A-14603
5	144.5	5.1	753	4	US-09-538-092-31
6	126	4.4	305	4	US-09-270-767-46577
7	126	4.4	2482	4	US-09-252-991A-16967
8	125.5	4.4	802	4	US-09-823-240A-2
9	123	4.3	884	6	5208144-8
10	123	4.3	884	6	5208144-8
11	120	4.2	542	4	US-09-949-016-6778
12	120	4.2	554	4	US-09-949-016-11687
13	120	4.2	950	4	US-09-252-991A-25927
14	114.5	4.0	1319	2	US-08-290-731C-2
15	114.5	4.0	1336	2	US-08-230-731C-6
16	113.5	4.0	2074	4	US-09-491-356C-9
17	111	3.9	806	4	US-09-902-540-14992
18	110.5	3.9	560	4	US-09-252-991A-25999
19	107.5	3.8	535	4	US-09-252-991A-17140
20	107	3.7	1183	3	US-09-134-001C-3530
21	106.5	3.7	328	4	US-09-252-991A-17729
22	106.5	3.7	574	4	US-09-270-767-45812
23	106	3.7	679	4	US-09-252-991A-18857
24	106	3.7	780	4	US-09-252-991A-32892
25	105.5	3.7	542	4	US-09-252-991A-32139
26	105.5	3.7	1706	4	US-09-252-991A-31760
27	105	3.7	416	4	US-09-252-991A-32660

28	105	3.7	561	4	US-09-252-991A-23080	Sequence 23080, A
29	105	3.7	781	4	US-09-949-016-9170	Sequence 9170, AP
30	104.5	3.7	582	3	US-08-906-865-3	Sequence 3, Appli
31	104.5	3.7	582	4	US-09-129-668-3	Sequence 3, Appli
32	104	3.6	406	4	US-09-252-991A-18696	Sequence 18696, A
33	104	3.6	955	4	US-09-949-016-8369	Sequence 8369, AP
34	103.5	3.6	248	4	US-09-252-991A-21872	Sequence 21872, A
35	103.5	3.6	798	3	US-09-150-460B-8	Sequence 8, Appli
36	103.5	3.6	875	3	US-09-150-460B-7	Sequence 7, Appli
37	103.5	3.6	1290	3	US-09-150-460B-6	Sequence 6, Appli
38	103	3.6	420	4	US-09-252-991A-20033	Sequence 20033, A
39	103	3.6	515	4	US-09-252-991A-32630	Sequence 32630, A
40	102.5	3.6	398	4	US-09-902-540-12305	Sequence 12305, A
41	102.5	3.6	1076	4	US-09-949-016-7421	Sequence 7421, AP
42	102	3.6	55	4	US-09-270-767-62317	Sequence 62317, A
43	102	3.6	809	4	US-09-252-991A-31759	Sequence 31759, A
44	102	3.6	3122	4	US-10-237-551-201	Sequence 201, App
45	102	3.6	3122	4	US-10-237-551-250	Sequence 250, App

ALIGNMENTS

RESULT 1
US-09-270-767-46703
; Sequence 46703, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46703
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46703

Query Match 7.4%; Score 212; DB 4; Length 561;
Best Local Similarity 29.2%; Pred No. 1.5e-11;
Matches 57; Conservative 37; Mismatches 75; Indels 26; Gaps 6;

QY 179 VVKYIETPLLCIDTKFDIRQWFLVTDNPLTIWFKESYLFSTQRF--LDKL-DSAI 235
DB 1 VVSKIYVDPLCIDGKCDLRVYVTSFDFLIIVYEEGIVRLATVKYDRHADNLWPC 60
QY 236 HLCNNAVQKYLKN-----DVGRSPLLPAHNMWSTTRFOEYLORQG---RGAVWG- 281
DB 61 HLCNYSINKYSHSYIRSSDAQDEVDG-----HKWTLSSALLRHKLQSCDTRQLMLNI 112
QY 282 -SVIYSPMKKAIHAHMKVAQDHVEPRKNSPELYGADPVLGRDFRPMLEIETNSPTMHPST 340
DB 113 EDIIKAVILACASIIISACRMFVPGNCPCLYGFDFILIDNAHKPMLLEINLSPSMGVDS 172
QY 341 PVTAQLCAQVQEDTI 355
DB 173 PLDTKVKSCLMADLL 187

RESULT 2
US-09-270-767-62411
; Sequence 62411, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517

```
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 62411  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-62411
```

Query Match
Best Local Similarity 5.4%; Score 154.5; DB 4; Length 260;
Matches 31.7%; Pred.No.1.7e-06;
Mismatches 40; Conservative 18; Mismatches 43; Indels 25; Gaps 3;

QY	125	TSVNPQTDI-----DGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPL	172
DG	136	TYNLSTEVKEFAAYYTAAAGLDNDHWIIKPMNLAARGLDTHITDNKIQLVRPAT----	191
QY	173	SRDKNWVVQKYETPFLIC-----DTKFDIRQWFVTWNPLTIWFYKSYLRFSTQR	226
DG	192	--GPKTAKXYIERPVLFPSRQEVEGSVKFDIRVILLKSVKPDKATHRKKFFLRPANHP	248
QY	227	SLDKLD 232	
DG	249	TLDHFDF 254	

RESULT 3

US-09-270-767-46787
Sequence 46787, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 46787
LENGTH: 507
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-46787

Query Match
Best Local Similarity 5.4%; Score 154.5; DB 4; Length 507;
Matches 31.7%; Pred.No.5e-06;
Mismatches 40; Conservative 18; Mismatches 43; Indels 25; Gaps 3;

QY	125	TSVNPQTDI-----DGLRNIIWKPAKSRGRDIVCMDRVVEEILELAADHPL	172
DG	383	TYNLSTEVKEFAAYYTAAAGLDNDHWIIKPMNLAARGLDTHITDNKIQLVRPAT----	438
QY	173	SRDKNWVVQKYETPFLIC-----DTKFDIRQWFVTWNPLTIWFYKSYLRFSTQR	226
DG	439	--GPKTAKXYIERPVLFPSRQEVEGSVKFDIRVILLKSVKPDKATHRKKFFLRPANHP	495
QY	227	SLDKLD 232	
DG	496	TLDHFDF 501	

RESULT 4

US-09-248-796A-14603
Sequence 14603, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409

Db 643 NI-KEQHSITNDVFLAANNRNLNFOPLNFAFTYGVDFLIDNSYEVKLEINAFDFK 701
Qy 338 PSTVPTAQLCAQVEDTIKAV 359
Db 702 QTGKDLKLNLDLFDFTVKYCV 723

RESULT 6

US-09-270-767-46577
; Sequence 46577, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46577
; LENGTH: 305
; TYPE: PRT
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-46577

Query Match 4.4%; Score 126; DB 4; Length 305;
Best Local Similarity 23.6%; Pred. No. 0.0013;
Matches 42; Conservative 36; Mismatches 62; Indels 38; Gaps 8;

Qy 180 VOKIETPLICDPKFDIOWFLVDNPLTIMFYKESYL-RFTQRFSLDKLDSAIHLC 238
Db 1 VQEFVQRPYLVDGHKFDIGVTIVTISVNLKRVIIYTGDLFRYCPVKYHPFDAEN----- 55
Qy 239 NNAVQKYLKNDVGRSPLPA-----HNMM---TSTRFQEYLQRQGR--CAVWGSGVI 284
Db 56 ---VDKIIVGD---DYLPTWEVPSLRKYNRFGSGMRTVFEAYVRDQKDPQAIWQVE 108
Qy 285 Y-----PSMKKAIAMKVAQDHVEPRKNSFELYGADFVLGRDPRFWLIEINSSPTM 336
Db 109 HIVRTTIAAKEKDIVNLSRVTH-----NFFDLMRFDLFDLKVFLMEANMSPNL 161

RESULT 7

US-09-252-991A-16967
; Sequence 16967, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16967
; LENGTH: 2482
; TYPE: PRT
; ORGANISM: *Pseudomonas aeruginosa*
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1766)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-16967

Query Match 4.4%; Score 126; DB 4; Length 2482;
Best Local Similarity 19.7%; Pred. No. 0.038;
Matches 122; Conservative 79; Mismatches 207; Indels 210; Gaps 29;

Qy 71 EHEDIDT-----SADAVEDLTEAEWEDLTQOYVSLVHGDAFISNSRNYFSOCOALLN 122
Db 308 DHTVDAERARLAGPRDAMRSVVGALCBELVRIKDSL-----DLFVRSDRGHPSELDALLA 364
Qy 123 RITSVNPQTDIDG-----LRNIWIIKPAKSRGR--DIVCMKDRVBEILELAA----- 167
Db 365 PLQAIADTLAVLFGQPRKVIDQLDVIHALAQCRREPSDAILMDVAGALLYVEATLAGM 424
Qy 168 ---ADHPLSRDNKVVQKIYET-PLLICDTKFDIQQ-----WFLVTDMNPLTIWFYKE- 216
Db 425 AGPGDERNSEESRLPTTVDVAQHQLVKEARNLEQAKDAIIBFIASQMNHEHLARVPEL 484
Qy 217 -SYLRFSTQRFSLDKLDSAIHLCLNNAVQKYLKNDVGRSPLLPANMMWTS-----TRF 267
Db 485 LTVRGGLAMIPLEAATLLETCTNYIQEQL-----LARK-----AVPDWQSLDTLADAITSV 537
Qy 268 QEYLQRQGRGAVWGSGVIYPSMKKAIAMKVAQDHVEPRKNSFELYGADFVLGRDPRFWL 327
Db 538 EYLLERL-----SEDHASQSDLI---LDVAEDSLAN-----LGYYTLKP-- 572
Qy 328 IEINS-----SPTMHPSTPVTALCAQ-----VOEDTIKVA 358
Db 573 ---NSSAPABPGLSGPAATESPAAPERPEAVVEVAETABEQPPADTAPAEAREDAFOLA 629
Qy 359 VDRSCDIG-----NPELLWRQPV---VEPPP-----F 382
Db 630 SDNWTILGEVAPDAGPSPDLALDLDDSAEVPFALPEVVEESGQPSQTPAPARSLLDF 689
Qy 383 SGSDLCVAGSVRRARRQVLPVCLNKASALLDAQPLKARGPSAMPD---PAQ----- 432
Db 690 SLDEIDLGLD-----LPA-----DAAP--ASGPAALADWSLPEQWGLGDDL 729
Qy 433 -GPSPALQDLDLKE-----EKGLPLALLAPLRGAESG---GAAQP-----TRT 474
Db 730 AQPTQAGETLIDLSLEBPALSFDAPLESLEPLALEPFDGSAEQELVLDALDPLDLVALP 789
Qy 475 KAAGKV-----ELPACPCRHVDQAPNTGVPVQAQPAKSWDPNOLNAHPL 518
Db 790 EAEGEVSAWEGSLEELDLSDDLPEVLPEAEAEAPPAEALASEAPALSLAEVMAAPV 849
Qy 519 EPVLRLKTAEGALRPPP 536
Db 850 QPINPPAQNPVSLPPP 867

RESULT 8

US-09-823-240A-2
; Sequence 2, Application US/09823240A
; Patent No. 6716597
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M00656.70064.US
; CURRENT APPLICATION NUMBER: US/09/823,240A
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: *Mus musculus*
US-09-823-240A-2

Query Match 4.4%; Score 125.5; DB 4; Length 802;
Best Local Similarity 24.7%; Pred. No. 0.0068;
Matches 78; Conservative 15; Mismatches 110; Indels 113; Gaps 15;

```
QY 264 STRFOYLORQGRGAVWGSVIYPSMKKAIAHAMKVADHVEPRKNSFELYGADFVLGRDP 323
Db 322 ATREATSL-----GSAFHP-----VLPHYATV-----PRP-----LNKNS 351
QY 324 RPLWLEINSSPTMHPST-PVTAOLCAQVQEDTTIKVAVDRSCDIGNFELLWRQP--VVEPP 380
Db 352 RP-----SSPVNTSPSQPPAAKSCA-----WPTSNSPLPPSPPINMISSP 391
QY 381 PFGSDLCVAGSVRRARRQVLPVC-----NLKASASLLDDAOPLKARGPS 425
Db 392 PGKAT-----GPREVLFCVSSVPVQMPSPPTAPNGSLDSVTYVPSPPTSGFA 440
QY 426 AMPDDPAQGPSPALQRDGLGKEKGPLALLAPLRGAESGGAQP-----471
Db 441 APPPPPPPPPP-----PPPLPLPPPLPLPLASLSHCGSQASPPPGTPLASTSSKP 491
QY 472 -----TRTKAGKVLPACPCRHVDSQAPNTGVPVA-OPAKSWDPNQL-----NAHPLEVP 521
Db 492 SVLPSPSAGAPASAEPLNP-ELGDSASEPGLQAASQPAESPTQGLVLGPPAPPPPPP 550
QY 522 LRGLKTAEGALRPPPG 537
Db 551 LPSPRAYASALPPPG 566

RESULT 9
5208144-8
; Patent No. 5208144
; APPLICANT: SMITH, JOHN A.; RAYCHOWDHURY, RAKTIMA; NILES, JOHN L.
; TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
; CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 42
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/396,697
; FILING DATE: 22-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 313,682
; FILING DATE: 22-FEB-1989
; APPLICATION NUMBER: 235,211
; FILING DATE: 23-AUG-1988
; SEQ ID NO: 8:
; LENGTH: 884
5208144-8
```

```
Query Match 4.3%; Score 123; DB 6; Length 884;
Best Local Similarity 19.7%; Pred. No. 0.014;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;

QY 66 YLGQLEH-EDITSADAVEDLTEAEWEDLTQQYYS-LVHGDAFISNRYFSQCQALLNR 123
Db 328 YLEEEHITQIDYDWD-----PEHIGLSVYVYTVLAQGSQFGAIKRAYIPNFESGSN- 379

QY 124 ITSVNPQTDID-GLRNIWIKP---AAKSRGRDIVCMRVEEILEAAADHPLSRDNKW 179
Db 380 ----NPIREVDLGLK--YLMQPDGLAVDWGRHIVWSANDANSORIEVATLD---GRYRKWL 430

QY 180 VQKIETPLLICDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLNC 239
Db 431 ITTOLDQPAAIA-----VNP-----KLGIMFWTDGQKPKIESA-----464

QY 240 NAVQKYLKNDVGRSPLLPANHMMWTSTRFOYLQORQGVAVGWSVIYPSMKKAIAHAMK-- 297
Db 465 -----WMNGEHSVLVSENGLWPNGLSIDYLNDD-----RVYWSDSKEDVIEAIKYD 511

QY 298 -----VAQDHVEP-----RKNSFELYGADFVLGRDPRPMLIEI 330
Db 512 GTDRRLIINEAMKPFSLDIFEDKLYWAKEGEVWRQNKFKENKEKVL--VVPNPLTVQ 569

QY 331 N-----SSPTMHPSTPVTAQI-----CAQVQEDTIKVAVDRSCDIGNFELLWRQ 374
Db 570 RIFHQLRYNQSVPNCKQVCSHLCLLRPGYSCACPGQSDFTVGTSTVQCDAAAS-EL---- 624

QY 375 PVVEPPP-----FSGDLCVAGSVRRARRQVLP-----403
Db 625 PVTPPPPCRCMGGNGCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 678

QY 404 -----VCNLKASASLLDDAOPLKARGPSAMPDPAQG-----433
Db 679 LTFVIVIIVGALVVLGFLHYRKTGSLPTLP-KLPSSLAKPSENGNGVTFRSGADVNM 737

QY 434 -----PPSPALQDILGKE-----EKG-LPLALLAPLRGAESGGAQPTRTKAAGKVE 481
```

```
Db 625 PVTPPPPCRCMGGNGCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 678
QY 404 -----VCNLKASASLLDDAOPLKARGPSAMPDPAQG-----433
Db 679 LTFVIVIIVGALVVLGFLHYRKTGSLPTLP-KLPSSLAKPSENGNGVTFRSGADVNM 737
QY 434 -----PPSPALQDILGKE-----EKG-LPLALLAPLRGAESGGAQPTRTKAAGKVE 481
Db 738 DIGVSPFGPGETIIDRSMAHNEHFVMEVGKQVIFENPMYAAKDN-----TSKVA 786
QY 482 IPACPCRHVDSQAPNTGVPVAOPAKSWDPNQLNAHPLEVPLGLKTAEGALRPPPGKGS 541
Db 787 LAV-----QGPSTGAQVTPENV--ENQNYGRPIDP-----SEIVPEPKPASPGA 829

RESULT 10
5208144-8
; Patent No. 5208144
; APPLICANT: SMITH, JOHN A.; RAYCHOWDHURY, RAKTIMA; NILES, JOHN L.
; TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
; CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 42
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/396,697
; FILING DATE: 22-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 313,682
; FILING DATE: 22-FEB-1989
; APPLICATION NUMBER: 235,211
; FILING DATE: 23-AUG-1988
; SEQ ID NO: 8:
; LENGTH: 884
5208144-8
```

```
Query Match 4.3%; Score 123; DB 6; Length 884;
Best Local Similarity 19.7%; Pred. No. 0.014;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;

QY 66 YLGQLEH-EDITSADAVEDLTEAEWEDLTQQYYS-LVHGDAFISNRYFSQCQALLNR 123
Db 328 YLEEEHITQIDYDWD-----PEHIGLSVYVYTVLAQGSQFGAIKRAYIPNFESGSN- 379

QY 124 ITSVNPQTDID-GLRNIWIKP---AAKSRGRDIVCMRVEEILEAAADHPLSRDNKW 179
Db 380 ----NPIREVDLGLK--YLMQPDGLAVDWGRHIVWSANDANSORIEVATLD---GRYRKWL 430

QY 180 VQKIETPLLICDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLNC 239
Db 431 ITTOLDQPAAIA-----VNP-----KLGIMFWTDGQKPKIESA-----464

QY 240 NAVQKYLKNDVGRSPLLPANHMMWTSTRFOYLQORQGVAVGWSVIYPSMKKAIAHAMK-- 297
Db 465 -----WMNGEHSVLVSENGLWPNGLSIDYLNDD-----RVYWSDSKEDVIEAIKYD 511

QY 298 -----VAQDHVEP-----RKNSFELYGADFVLGRDPRPMLIEI 330
Db 512 GTDRRLIINEAMKPFSLDIFEDKLYWAKEGEVWRQNKFKENKEKVL--VVPNPLTVQ 569

QY 331 N-----SSPTMHPSTPVTAQI-----CAQVQEDTIKVAVDRSCDIGNFELLWRQ 374
Db 570 RIFHQLRYNQSVPNCKQVCSHLCLLRPGYSCACPGQSDFTVGTSTVQCDAAAS-EL---- 624

QY 375 PVVEPPP-----FSGDLCVAGSVRRARRQVLP-----403
Db 625 PVTPPPPCRCMGGNGCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 678

QY 404 -----VCNLKASASLLDDAOPLKARGPSAMPDPAQG-----433
Db 679 LTFVIVIIVGALVVLGFLHYRKTGSLPTLP-KLPSSLAKPSENGNGVTFRSGADVNM 737

QY 434 -----PPSPALQDILGKE-----EKG-LPLALLAPLRGAESGGAQPTRTKAAGKVE 481
```

Db 738 DIGVSPFGPETIIDRSAMNHFVMEVGKQVPVFENPMYAAKDN-----TSKVA 786
Qy 482 LPAPCPCRHVSQAQNTGVPVQAQPAKSWDPNQLNAHLEPVLRLGKLTAEAGLRPPPGKGKS 541
Db 787 LAV-----QGPSTGAQVTPENV--ENQYGRPIDP-----SEIVPEPKPASPFGA 829

RESULT 11

US-09-949-016-6778
; Sequence 6778, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6778
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6778

Query Match 4.2%; Score 120; DB 4; Length 542;
Best Local Similarity 22.5%; Pred. No. 0.012;
Matches 75; Conservative 38; Mismatches 124; Indels 96; Gaps 18;

Qy 254 PLLPAHNMWTSRFEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 313
Db 233 PLLPGNDV-----GRSS-YGAM--QVKQVFDYAYIVLASHVSPARSYPNR 275
Qy 314 GADPVLGR-----DFRPWLEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCD 364
Db 276 DAESTLGRILIKVTQEVIDYRRWIKWKWSKA-HPSGMDSRI--KIKE---RIATCNQEQ 329
Qy 365 IGNFELLWRQPVVEPP-----PFGSDLCVAGSVRRARR-----QVLPVCNLKA- 409
Db 330 TQN-----REP--ESPYGQRLTLSSLSPQLSSGSSASSVSSLSGSDVSDTTPCTTPSV 382
Qy 410 -SASLLDAQPLKARGSPAMPDPAQPPSPALORDLGLKEEGLPLALLAPLRGAA-----E 464
Db 383 YQFSLOQAPAPLMAGLTALPMPS-GKPOPTTSRTLTMTNNQTRFTTIPPTLGVAPVPCR 441
Qy 465 SGGAAPQTRTKAAGKVELPACPCRHVDSQAPNTGVPVQAQPAKSWDPNQLNAHLEPVLRG 524
Db 442 QAGVEGTASLKAHVHMSSPAIP-----SASPN---PLSSP-----HLVHKQHNG 482
Qy 525 LK-----TAEGALRPPPGKGK 540
Db 483 MKLSMKGSHGHTGGGYSSVSGGVRRPPVGNRG 515

RESULT 12

US-09-949-016-11687
; Sequence 11687, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11687
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11687

Query Match 4.2%; Score 120; DB 4; Length 554;
Best Local Similarity 22.5%; Pred. No. 0.013;
Matches 75; Conservative 38; Mismatches 124; Indels 96; Gaps 18;

Qy 254 PLLPAHNMWTSRFEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFELY 313
Db 245 PLLPGNDV-----GRSS-YGAM--QVKQVFDYAYIVLASHVSPARSYPNR 287
Qy 314 GADPVLGR-----DFRPWLEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCD 364
Db 288 DAESTLGRILIKVTQEVIDYRRWIKWKWSKA-HPSGMDSRI--KIKE---RIATCNQEQ 341
Qy 365 IGNFELLWRQPVVEPP-----PFGSDLCVAGSVRRARR-----QVLPVCNLKA- 409
Db 342 TQN-----REP--ESPYGQRLTLSSLSPQLSSGSSASSVSSLSGSDVSDTTPCTTPSV 394
Qy 410 -SASLLDAQPLKARGSPAMPDPAQPPSPALORDLGLKEEGLPLALLAPLRGAA-----E 464
Db 395 YQFSLOQAPAPLMAGLTALPMPS-GKPOPTTSRTLTMTNNQTRFTTIPPTLGVAPVPCR 453
Qy 465 SGGAAPQTRTKAAGKVELPACPCRHVDSQAPNTGVPVQAQPAKSWDPNQLNAHLEPVLRG 524
Db 454 QAGVEGTASLKAHVHMSSPAIP-----SASPN---PLSSP-----HLVHKQHNG 494
Qy 525 LK-----TAEGALRPPPGKGK 540
Db 495 MKLSMKGSHGHTGGGYSSVSGGVRRPPVGNRG 527

RESULT 13

US-09-252-991A-25927
; Sequence 25927, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25927
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25927

Query Match 4.2%; Score 120; DB 4; Length 950;
Best Local Similarity 25.1%; Pred. No. 0.031;
Matches 53; Conservative 17; Mismatches 63; Indels 78; Gaps 9;

Qy 361 RSCDIGNFELLWRQPVVEPPPF---SGSDLCVAGSVR-----RARRQVLPVCNLKA 409
Db 728 RTADVS-----RQPLRQPPALLQRRGVDLASLRPPGPPGAARALLSPACRRGR 781
Qy 410 SASLLDAQP-----LKARGPSAM-----PDPAQG-----PPSPALORDLGLK 446

Db 782 ATGLGQPGGRGRLSRCPGAAATAAGPGCTAHARGGGDGLRRRHPPAPARPRGPGLA 841
QY 447 EEKGLPLALLALRGAESGGAQPTRTKAAGKVELPACPCRHVDSOAPNTGVPVAPAK 506
Db 842 P-----APAGGADPGAAQGRWPA-----PAARROR 869
QY 507 SWDPNQLNAHPLRPVLRLGLKTABGALRPPPG 537
Db 870 TWTTRRRS-----LRTLERPAGVPAPPG 894

RESULT 14
US-08-290-731C-2
; Sequence 2, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (mSOS) GENE,
; TITLE OF INVENTION: AND mSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PLO921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-2

Query Match 4.0%; Score 114.5; DB 2; Length 1319;
Best Local Similarity 19.2%; Pred. No. 0.18;
Matches 118; Conservative 82; Mismatches 204; Indels 209; Gaps 31;

QY 62 VCOAYLGQLEHEDIDTSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQOALL 121
Db 692 VCRHW---VEHFYDFERDA---DLLQR-----MEEFITGVKGAM---KKWVESITKII 737
QY 122 NR-----ITSVNPQTDIDGLRLNIWIKPAKSGRGRDIVCMGRVVEETLELAAD 169
Db 738 QRKKIARDNGPGHNITFQSSPPTVE-----WHLSRPGHETFDLLTHPIETARQLTLE 792
QY 170 HPLSR-----DNKVVQV-KYIETPLICDTKFDIQRWFLVDWNPLTIWPK----- 215

Db 793 SLDYRAVQPSLVGSVWTKEDKEINSPLL-----KMRHTTNLTLMFEKICIVET 842
QY 216 -----ESYLRF-----STORESLDK-----LD 232
Db 843 ENLEERVAVVSRIIIEILQVFQELNNFNGVLVVSAMNSSFVRLDHTTFQIISRQKKILE 902
QY 233 SAHLCNNAVOKYL-KNDVGRSPLLPAHNMW-----TSTRFQEYLQROGRGAVMGSVIY 285
Db 903 EAHELSEHDYKYLAKLRSINPCVPFFGILYLTNLTKEGNEPVLRRHGK-----ELIN 957
QY 286 PNMKKAIA-----HAMKVAQD-----HVEPRKNSFELYGADFVLGRDPRFW 326
Db 958 FSKRRRVAEITGEIQYQYQYQYCLRVDPDKRFFENLNPNGNSMEKEFTDYLFNKS----- 1013
QY 327 LIEIN-----SPTMEHPT--PVTAQLCAQVQEDTIKVAVDR----- 361
Db 1014 -LEIBPRHPKPLRPFPKKYSYPLKSGVVRPSNRPOTGMTRHPTLQOEPKIKISYSRIPSE 1072
QY 362 ---SCDIGNFELLWRQPVVPPFSG-----SDLCVAGSVRRARQVLFVCLNKASASLL 414
Db 1073 TESTASAPNSP---RTPLT-PPPASGTSNTDVC---SVFSDSHSASPFHSRSASVSSI 1124
QY 415 DAQPLKARQPSAMPDPAQPPSPALQORDLGLKEKGLPLALLAPLRGAESGGAQPTRT 474
Db 1125 SL---SKGTDEVPVP---PPVPPRRR-----PESAPAES-----SPSKI 1157
QY 475 KAAGKVELPACPCRHVDSOAPNTGVPVAPAKSWDPNQLNAHPLRPVLRLGLKTAEQ--- 530
Db 1158 MSKHLDSPPAIPROPTSKAYSISDRTSISDPPE--SPPLLPPRPVPTPDVFSSS 1215
QY 531 --ALRPPPGGKGS 541
Db 1216 PLHLQPPPLGKKS 1228

RESULT 15
US-08-290-731C-6
; Sequence 6, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (mSOS) GENE,
; TITLE OF INVENTION: AND mSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PLO921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-2

Search completed: September 17, 2005, 08:52:07
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 08:48:28 ; Search time 169 Seconds

(without alignments)
1296.247 Million cell updates/sec

Title: US-10-615-659-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTAEGALRPPGGKGS 541

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1812044 seqs, 40492589 residues

Total number of hits satisfying chosen parameters: 1812044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20:	/cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	541	16	US-10-615-659-2
2	2854	100.0	541	16	US-10-635-977-2
3	2843	99.6	541	16	US-10-615-659-13
4	2843	99.6	541	16	US-10-635-977-13
5	1560	54.7	293	16	US-10-615-659-14
6	1560	54.7	293	16	US-10-635-977-14
7	1299	45.5	242	16	US-10-615-659-4
8	1299	45.5	242	16	US-10-635-977-4
9	1013	35.5	488	15	US-10-275-959A-5
10	969.5	34.0	362	9	US-09-925-298-617
11	969.5	34.0	362	14	US-10-102-806-617

12	956.5	33.5	352	16	US-10-615-659-7	Sequence 7, Appli
13	956.5	33.5	352	16	US-10-635-977-7	Sequence 7, Appli
14	956.5	33.5	352	18	US-10-756-149-5267	Sequence 5267, Ap
15	944.5	33.1	326	15	US-10-108-260A-4209	Sequence 4209, Ap
16	865.5	30.3	292	16	US-10-615-659-5	Sequence 5, Appli
17	865.5	30.3	292	16	US-10-635-977-5	Sequence 5, Appli
18	375.5	13.2	566	18	US-10-450-763-36330	Sequence 36330, A
19	322	11.3	330	16	US-10-615-659-26	Sequence 26, Appl
20	322	11.3	330	16	US-10-635-977-26	Sequence 26, Appl
21	303.5	10.6	1226	18	US-10-756-149-5184	Sequence 5184, Ap
22	302	10.6	268	15	US-10-424-599-205823	Sequence 205823,
23	299.5	10.5	423	16	US-10-615-659-8	Sequence 8, Appli
24	299.5	10.5	423	16	US-10-635-977-8	Sequence 8, Appli
25	280.5	9.8	439	15	US-10-104-047-3658	Sequence 3658, Ap
26	274	9.6	49	16	US-10-615-659-21	Sequence 21, Appl
27	274	9.6	49	16	US-10-615-659-22	Sequence 22, Appl
28	274	9.6	49	16	US-10-635-977-21	Sequence 21, Appl
29	274	9.6	49	16	US-10-635-977-22	Sequence 22, Appl
30	271.5	9.5	524	9	US-09-864-761-38213	Sequence 38213, A
31	271.5	9.5	524	15	US-10-104-047-3371	Sequence 3371, A
32	259	9.1	379	16	US-10-615-659-6	Sequence 6, Appli
33	259	9.1	379	16	US-10-635-977-6	Sequence 6, Appli
34	258	8.9	92	15	US-10-424-599-262294	Sequence 262294,
35	254.5	8.9	377	15	US-10-210-130-118	Sequence 118, App
36	254.5	8.9	377	15	US-10-250-613-1	Sequence 1, Appli
37	237	8.3	42	16	US-10-615-659-24	Sequence 24, Appl
38	237	8.3	42	16	US-10-635-977-24	Sequence 24, Appl
39	207	7.3	698	15	US-10-108-260A-4312	Sequence 4312, Ap
40	207	7.3	767	18	US-10-450-763-41893	Sequence 41893, A
41	205.5	7.2	553	15	US-10-108-260A-2495	Sequence 2495, Ap
42	205.5	7.2	887	16	US-10-473-574-20	Sequence 20, Appl
43	186	6.5	663	16	US-10-322-281-133	Sequence 133, App
44	183	6.4	404	15	US-10-104-047-3621	Sequence 3621, Ap
45	178.5	6.3	237	18	US-10-450-763-41569	Sequence 41569, A

ALIGNMENTS

RESULT 1

US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

Query Match	100.0%	Score	2854	DB	16	Length	541
Best Local Similarity	100.0%	Pred. No.	3e-229				
Matches	541	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MASSILKVVVSHQSCSRSPRDQREAGSSDLSRQDAENAEAKRLGPGQVLDIAC	60				
Db	1	MASSILKVVVSHQSCSRSPRDQREAGSSDLSRQDAENAEAKRLGPGQVLDIAC	60				
Qy	61	KVQCAVYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYSLVHGDFTSNRNYFSQCOAL	120				
Db	61	KVQCAVYLGQLEHEDIDTSADAVEDLTEAEWEDLTQQYSLVHGDFTSNRNYFSQCOAL	120				
Qy	121	LNRTSVNPTQDIDGLRNWIIPAAKSRGRDIVCMRDVEEILELAAADHPLSRDNKVVV	180				

Db 121 LNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
QY 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN 240
Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN 240
QY 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
QY 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARQVLPVCNLKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARQVLPVCNLKASASLLDAQPLK 420
QY 421 ARGPSAMPDPAQPPSPALORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
QY 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGTAEAGLRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGTAEAGLRPPPGKG 540
QY 541 S 541
Db 541 S 541

RESULT 2
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 2854; DB 16; Length 541;
Best Local Similarity 100.0%; Pred. No. 3e-229;
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASSILKVVVSHQSCSRSSKPRDQREAGSDLSRQDAENAEAKRLGLPQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSKPRDQREAGSDLSRQDAENAEAKRLGLPQLVDIAC 60
QY 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSOCQAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSOCQAL 120
QY 121 LNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
Db 121 LNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
QY 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN 240

Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN 240
QY 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
QY 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARQVLPVCNLKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARQVLPVCNLKASASLLDAQPLK 420
QY 421 ARGPSAMPDPAQPPSPALORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
QY 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGTAEAGLRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGTAEAGLRPPPGKG 540
QY 541 S 541
Db 541 S 541

RESULT 3
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13

Query Match 99.6%; Score 2843; DB 16; Length 541;
Best Local Similarity 99.6%; Pred. No. 2.5e-228;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MASSILKVVVSHQSCSRSSKPRDQREAGSDLSRQDAENAEAKRLGLPQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSKPRDQREAGSDLSRQDAENAEAKRLGLPQLVDIAC 60
QY 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSOCQAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSOCQAL 120
QY 121 LNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
Db 121 LNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVV 180
QY 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN 240
Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLDSAIHLCNN 240
QY 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDRFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360

Db 301 DHVEPRKNSFELYGADFLVGRDFRPMLEINSSPTMHPSTPVTQAQVQEDTIKVAVD 360
Qy 361 RSDIGNFELLWRQPVVEPPFPSSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Db 361 RSDIGNFELLWRQPVVEPPFPSSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPLALLAPLRGAABSGGAAQPTTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPLALLAPLRGAABSGGAAQPTTRTKAAGKV 480
Qy 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKTAEGALRPPPGKG 540
Db 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKTAEGALRPPPGKG 540
Qy 541 S 541
Db 541 S 541

RESULT 4

US-10-635-977-13
; Sequence 13, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-13

Query Match 99.6%; Score 2843; DB 16; Length 541;
Best Local Similarity 99.6%; Pred. No. 2.5e-288;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MASSILKVVVSHQSCRSRSPKRDQREAEAGSSDLSSRQDAENAEAKLRLPGQLVDIAC 60
Db 1 MASSILKVVVSHQSCRSRSPKRDQREAEAGSSDLSSRQDAENAEAKLRLPGQLVDIAC 60
Qy 61 KVCQAYLGQLEHEDIDTSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCCAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCCAL 120
Qy 121 LNRIITSVNPQTDIGLRIWIIPKAASGRDIVCMRVVEIELEAAADHPLSRDNKVV 180
Db 121 LNRIITSVNPQTDIGLRIWIIPKAASGRDIVCMRVVEIELEAAADHPLSRDNKVV 180
Qy 181 QKIYETPLLCIDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSFORPSLDKLSAHLN 240
Db 181 QKIYETPLLCIDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSFORPSLDKLSAHLN 240
Qy 241 AVQKYLKNDVGRSPLLPAAHNMWTSRFEYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLPAAHNMWTSRFEYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQ 300
Qy 301 DHVEPRKNSFELYGADFLVGRDFRPMLEINSSPTMHPSTPVTQAQVQEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFLVGRDFRPMLEINSSPTMHPSTPVTQAQVQEDTIKVAVD 360
Qy 361 RSDIGNFELLWRQPVVEPPFPSSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420

Db 361 RSDIGNFELLWRQPVVEPPFPSSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPLALLAPLRGAABSGGAAQPTTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPLALLAPLRGAABSGGAAQPTTRTKAAGKV 480
Qy 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKTAEGALRPPPGKG 540
Db 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKTAEGALRPPPGKG 540
Qy 541 S 541
Db 541 S 541

RESULT 5

US-10-615-659-14
; Sequence 14, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-14

Query Match 54.7%; Score 1560; DB 16; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e-121;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 73 EDIDTSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCCALLNRITSVNPQTD 132
Db 1 EDIDTSADAVEDLTAEWEDLTQYYSLVHGDAFISNSRNYFSQCCALLNRITSVNPQTD 60
Qy 133 IDGLRNIWIIPKAASGRDIVCMRVVEIELEAAADHPLSRDNKVVQKIYETPLICD 192
Db 61 IDGLRNIWIIPKAASGRDIVCMRVVEIELEAAADHPLSRDNKVVQKIYETPLICD 120
Qy 193 TKEDIRQWFLVTDWNPPLTIWFYKESYLRFSFORPSLDKLSAHLNNAVQKYLKNDVGR 252
Db 121 TKEDIRQWFLVTDWNPPLTIWFYKESYLRFSFORPSLDKLSAHLNNAVQKYLKNDVGR 180
Qy 253 SPLLPAAHNMWTSRFEYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 312
Db 181 SPLLPAAHNMWTSRFEYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 240
Qy 313 YGADFVLGRDFRPMLEINSSPTMHPSTPVTQAQVQEDTIKVAVDRCSDI 365
Db 241 YGADFVLGRDFRPMLEINSSPTMHPSTPVTQAQVQEDTIKVAVDRCSDI 293

RESULT 6

US-10-635-977-14
; Sequence 14, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725

;
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-14

Query Match 54.7%; Score 1560; DB 16; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e-121;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 EDIDTSADAVEDTEAEWEDLTQOYSLVGHDAFISNRYFSQCCALLNRITSNVPQTD 132
DB 1 EDIDTSADAVEDTEAEWEDLTQOYSLVGHDAFISNRYFSQCCALLNRITSNVPQTD 60
QY 133 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILELAAADHPLSRDNKVVQKYIETPLLCD 192
DB 61 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILELAAADHPLSRDNKVVQKYIETPLLCD 120
QY 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
DB 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 180
QY 253 SPLLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
DB 181 SPLLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240
QY 313 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 365
DB 241 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 293

RESULT 7
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4

Query Match 45.5%; Score 1299; DB 16; Length 242;
Best Local Similarity 100.0%; Pred. No. 6.8e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILELAAADHPLSRDNKVVQKYIETPLLCD 192
DB 1 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILELAAADHPLSRDNKVVQKYIETPLLCD 60
QY 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
DB 61 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
QY 253 SPLLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
DB 121 SPLLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

QY 313 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 372
DB 181 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 240
QY 373 RQ 374
DB 241 RQ 242
RESULT 8
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-4

Query Match 45.5%; Score 1299; DB 16; Length 242;
Best Local Similarity 100.0%; Pred. No. 6.8e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILELAAADHPLSRDNKVVQKYIETPLLCD 192
DB 1 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILELAAADHPLSRDNKVVQKYIETPLLCD 60
QY 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 252
DB 61 TKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQFSLDKLDSAIHLCCNNAVQKYLKNDVGR 120
QY 253 SPLLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
DB 121 SPLLPAHNMTSTRFOEYLQQRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180
QY 313 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 372
DB 181 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDI 240
QY 373 RQ 374
DB 241 RQ 242

RESULT 9
US-10-275-595A-5
; Sequence 5, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: SAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Valda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil

; APPLICANT: BATRA, Sajeev
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No: 2156553CD1
US-10-275-595A-5

Query Match 35.5%; Score 1013; DB 15; Length 488;
Best Local Similarity 45.5%; Pred. No. 1.4e-75;
Matches 214; Conservative 68; Mismatches 138; Indels 50; Gaps 10;
Qy 74 DITSADAVEDLTEAEVEDLTQYYISLVHGDAFTNSRNTFSQCALNLRITSVNPQTDI 133
Db 2 DIDKDLAPLYLTPEGHSLFLQRYQVYVHGAELRLDVTQVRCEDILQQLQAVVPQIDM 61
Qy 134 DGLRNIIIPAAKSRGRDVCMDRVEILELAADHPLSRDNKVVQKYIETPLLICDT 193
Db 62 EGRNIIIVPGAKSRGRGIMCDHLEMLKLVNGPNVWKGKVVQKYIERPLLIFFT 121
Qy 194 KFDIROWFLVTWNPLTIWFYKESYLRFSTQRFSLDKLDSAIHLCNNAVQKLYKNDVGRS 253
Db 122 KFDLQWFLVTWNPLTIWFYRDSYIRFSTQPFSLKNLDSVHLNNSIQHLENSCHRH 181
Qy 254 PLPAHNNWTSTRFOEYLQGRGAVNGSVIYPSMKKIAIAHAMKVAQDHVPRKNSFELY 313
Db 182 PLLPPDNWSSQRFQAHLEQMGAPNASTIIVPGMDAVIHALQTSQDVTQCRKASFELY 241
Qy 314 GADFLVGRDPRPWLIEINSPHPSTVTAQLCAQVQEDTIKAVV---DRSCDIGNFE 369
Db 242 GADPVFGEDFQPLWIEINASTPMASTAVTARLCAGVQADTLRVVIDRMLDRNCDTGAFE 301
Qy 370 LLMRQPVVEPPFSGDLVCVAGSVRRARRQVLPVCNLKASASLLDAQPLKARGPSAMPD 429
Db 302 LIYQPAVEVQYVIRLLVGEFTIKK-----PMAMCH 334
Qy 430 PAQG--PPSPAL-ORDLGLKEEKLGLPALLAPLR--GAASGGAAQP---TRTKAAG-- 479
Db 335 RRMGVRPAVPLLTQSGEGEKDSIGIPTHRSASRKGTGARSIGHSEKDVSTATTSPAGKGK 394
Qy 480 ----VELPACPCRHV-DSQAPNT--GVYPAQPAKSWD--PNQLNAHLEP 520
Db 395 KGKAKRATALVCPNLWEWDAPSTRMGCIFTWTFSSGDRQPHHLNRLPLSP 444

RESULT 10
US-09-925-298-617
; Sequence 617, Application US/09925298
; Publication No. US20020039764A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103

; CURRENT APPLICATION NUMBER: US/09/925,298
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-298-617
Query Match 34.0%; Score 969.5; DB 9; Length 362;
Best Local Similarity 52.0%; Pred. No. 4e-72;
Matches 183; Conservative 58; Mismatches 92; Indels 19; Gaps 3;
Qy 49 RGLPGQLVDIACKVCOAYLQGLEHEDIDTSADAVEDLITEAEWEDLTQYYISLVHGDAFIS 108
Db 9 RGVPYQ-----LGPHGHRQ---GLEAPLYLTPEGWSLFLQRYQVYVHGAELR 53
Qy 109 NSRNYFSQCALNLRITSVNPQTDIDGLRNIIIPAAKSRGRDVCMDRVEILELA 168
Db 54 HLDVTQVRCEDILQQLQAVVPQIDMEGRNIIIVPGAKSRGRGIMCDHLEMLKLVNG 113
Qy 169 DPLSRDNKVVQKYIETPLLICDTKFDIROWFLVTWNPLTIWFYKESYLRFSTQRFSL 228
Db 114 NPVMDKGRVVQKYIERPLLIETGTRFDLQWFLVTWNPLTIWFYRDSYIRFSTQRFSL 173
Qy 229 DKLDSAIHLCNNAVQKLYKNDVGRSPLPAHNNWTSTRFOEYLQGRGAVNGSVIYPSM 288
Db 174 KNLDSVHLNNSIQHLENSCHRHPLLPDNNWSSQRFQAHLEQMGAPNASTIIVPGM 233
Qy 289 KKAIAHAMKVAQDHVPRKNSFELYGADFVLGRDPRPWLIEINSSPTMHSTPVTALCA 348
Db 234 KDAVIHALQTSQDVTQCRKASFELYGADFVFGEDFQPLWIEINASTPMASTAVTARLCA 293
Qy 349 QVQEDTIKAVV---DRSCDIGNFE LLMRQPVVEPPFSGDLVCVAGSVRR 396
Db 294 GVQADTLRVVIDRXLDRNCDTGAFELIYKQPAVEVQYVIRLLVGEFTIKK 345

RESULT 11
US-10-102-806-617
; Sequence 617, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

This Page Blank (uspto)